TOWARDS UNDERSTANDING TENOFOVIR DISPOSITION: MALDI-MS IMAGING AND PHARMACOGENETICS

By Dominique B. Figueroa

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, MD

November 2016

Abstract

The emergence of HIV pre-exposure prophylaxis, applying antiretroviral drugs toward prevention of infection by human immunodeficiency virus in uninfected individuals, reveals a promising avenue toward decreasing the number of annual new HIV infections worldwide. The nucleotide reverse transcriptase inhibitor tenofovir, along with the nucleoside reverse transcriptase inhibitor emtricitabine, currently make up the FDA-approved regimen for pre-exposure prophylaxis. Intracellular phosphorylation of both tenofovir and emtricitabine is necessary to generate nucleotide triphosphate analogs that would competitively inhibit HIV reverse transcriptase.

Kinases demonstrated to phosphorylate tenofovir to the intermediate tenofovir monophosphate and then to the active anabolite tenofovir monophosphate have been shown to be expressed in HIV-susceptible tissues. Adenylate kinase 2 phosphorylates tenofovir to tenofovir monophosphate in peripheral blood mononuclear cells, vaginal, and colorectal tissues. In colorectal tissues, creatine kinase, muscle was found to phosphorylate tenofovir monophosphate to tenofovir diphosphate. The two kinases pyruvate kinase, muscle and pyruvate kinase, liver and red blood cell were shown to yield tenofovir diphosphate in peripheral blood mononuclear cells and vaginal tissue. These four kinases were sequenced for genetic variants that could affect tenofovir activation. In a clinical study spread across the three geographic regions Bangkok, Thailand, Cape Town, South Africa, and New York City, USA, 103 previously unreported variants across the kinases *AK2, CKM, PKM* and *PKLR* were detected. Of 505 individuals, 19 individuals (3.7% frequency) were found to be carrying variants that were predicted to be deleterious and probably damaging.

In emtricitabine activation, we put forward four candidate kinases that would yield the nucleotide triphosphate analog emtricitabine triphosphate. Emtricitabine requires three phosphorylation steps to become pharmacologically active. Deoxycytidine kinase and thymidine kinase 1 are proposed to phosphorylate emtricitabine to emtricitabine-monophosphate. To generate emtricitabine-diphosphate, cytidine monophosphate kinase 1 is suggested to phosphorylate emtricitabine-monophosphate. To form the competitive HIV reverse transcriptase inhibitor emtricitabine triphosphate, phosphoglycerate kinase 1 was identified as a candidate kinase. In a clinical study, we identified 17 previously unreported genetic variants of these kinases.

Additionally, to further understand tenofovir and emtricitabine disposition in tissue, a matrix-assisted laser desorption/ionization coupled to mass spectrometry method was developed. This method aims to image tenofovir, emtricitabine, and the corresponding phosphorylated anabolites of each drug in an effort to better understand drug distribution in HIV susceptible tissues.

Taken together, development of the imaging technology and application of next-generation sequencing to identify variants in kinases proposed to activate nucleos(t)ide reverse transcriptase inhibitors aim to allow for explanations of inter-individual variability of pre-exposure prophylaxis efficacy.

Doctoral advisor:	Dr. Namandjé N. Bumpus
Thesis readers:	Dr. Namandjé N. Bumpus
	Dr. Amanda M. Brown
Thesis committee:	Dr. Namandjé N. Bumpus
	Dr. Amanda M. Brown
	Dr. Craig W. Hendrix
	Dr. James T. Stivers

Acknowledgements

I would like to thank to my advisor Dr. Namandjé N. Bumpus for her support during my study and research. Her feedback helped me during researching and writing this thesis.

Besides Dr. Bumpus, I thank the rest of my thesis committee: Dr. Amanda M. Brown, Dr. Craig W. Hendrix, and Dr. James T. Stivers. I thank Dr. Brown for agreeing to be my thesis reader.

Thanks also goes to my labmates in the Bumpus Lab: Dr. Julie Lade, Dr. Phil Cox, Carley Heck, and Dr. Erin Madeen.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

I. HIV pre-exposure prophylaxis aims to reduce HIV incidence

Published in 2016, The World Health Organization Global AIDS Update reports over 36.7 million people worldwide living with human immunodeficiency virus (HIV) [1]. This takes into account the number of people living with HIV on antiretroviral therapy, however it also includes the up to 2.4 million new HIV infections reported in 2015 alone. Over the past five years, the number of new HIV infections has not appreciably declined, remaining static around 2 million new infections since 2010 worldwide [1]. In specific regions such as Eastern Europe and central Asia, this number has even doubled from 100,000 diagnoses in 2010 to 200,000 new diagnoses in 2015 [1]. Given this alarming and steady level of new infections, measures toward reducing the incidence of HIV infection are under active investigation. As the availability of antiretroviral therapies to HIV infected individuals has increased, the prevalence of HIV has as well. The number of HIV-infected individuals on antiretroviral therapy has more than doubled from 7.5 million people to 17.0 million since 2010 [1]. If the number of new infections per year does not decline, the prevalence of HIV worldwide will continue to rise.

Interventions to reduce the incidence of HIV infection have included behavioral, surgical, and pharmacological approaches such as increased condom use [2], voluntary medical male circumcision [3, 4, 5], and prevention of mother-to-child transmission of HIV [6, 7] respectively. A pharmacological approach, prevention of mother-to-child transmission, implements antiretroviral drugs that have been previously used to treat individuals who have already been infected with HIV [6]. In an example employing this

strategy, the Breastfeeding, Antiretrovirals, and Nutrition study used antiretrovirals such as nevirapine, a non-nucleoside reverse transcriptase inhibitor, zidovudine, and lamivudine, both nucleoside reverse transcriptase inhibitors. These pharmacological agents were applied toward combatting infection of HIV in HIV-uninfected infants [6]. This study showed a 74% protection rate in infants who had been administered nevirapine and a 53% protection rate in infants whose mothers were on antiretroviral therapy [6].

Administration of antiretrovirals for use against HIV infection has since been able to provide protection as post-exposure prophylaxis [8], as secondary prevention when given to HIV infected individuals [9], and as pre-exposure prophylaxis (PrEP) [10, 9]. Along with rates of protection from HIV infection in mother-to-child transmission studies, the rationale behind further investigation of PrEP lies within animal model studies. Several studies using macaque models and the antiretroviral nucleos(t)ide reverse transcriptase inhibitors tenofovir and emtricitabine indicate protection from HIV infection when either of these drugs are administered [11, 12, 13]. In one study, administration of either tenofovir alone or a combination of tenofovir and emtricitabine showed complete protection of macaques against simian-human immunodeficiency virus [12]. Supported by successful protection outcomes from mother-to-child transmission and macaque studies, PrEP was further investigated using tenofovir and emtricitabine in clinical trials that directly tested its efficacy.

II. Tenofovir and emtricitabine are approved for use as PrEP

A form of chemoprophylaxis, PrEP is a strategy for HIV prevention in which an individual who is not infected with HIV is administered an oral or topical formulation of

an anti-HIV drug to protect themselves against HIV infection [14]. Currently, the only FDA-approved formulation of PrEP is Truvada ®, an oral coformulation of the nucleoside/nucleotide reverse transcriptase inhibitors tenofovir and emtricitabine [15, 16].

Tenofovir and emtricitabine both mimic nucleosides/nucleotides that must be phosphorylated intracellularly to form their active metabolites [17, 18]. These phosphorylated metabolites are nucleoside/nucleotide triphosphate analogs that competitively inhibit HIV reverse transcriptase. Tenofovir is phosphorylated twice inside the cell to form tenofovir diphosphate [17], while emtricitabine is phosphorylated intracellularly three times to form emtricitabine triphosphate [18].

Prior to its approval for use as PrEP in 2012, four clinical trials investigated tenofovir and emtricitabine. Testing either oral or topical administrations of drug, these studies were able to demonstrate protection from HIV infection. The largest study, monitoring 4,758 couples in Kenya and Uganda, showed a 75% rate of protection from HIV infection in subjects assigned to taking either tenofovir or tenofovir in conjunction with emtricitabine [19]. All couples participating in this study, known as Partners PrEP, were sero-discordant, with one partner infected with HIV and the other an HIV-uninfected individual. The remaining three successful clinical trials examined populations of men who have sex with men in the iPrEx study [20], heterosexual men and women in the Botswana TDF2 study [21], and injection drug users in the Bangkok Tenofovir study [22, 23]. These studies showed 44%, 62%, and 49% reduction in HIV incidence respectively [20, 21, 23]. Taking adherence, measured by detection of tenofovir in participant plasma, into account, these statistics show greater rates of protection from

HIV infection. The Partners PrEP study exhibited 90% protection, iPrEx 92%, the Botswana TDF2 study 78%, and the Bangkok Tenofovir study 74% when examining participants with detectable tenofovir plasma levels [19, 20, 21, 23]. The increased rates of protection from HIV infection generated by these large-scale clinical trials across diverse populations resulted in the FDA's approval of Truvada ® as PrEP in July 2012.

Despite taking adherence into account when examining rates of PrEP efficacy across the four clinical trials, complete protection against HIV infection was not observed. Indeed, two clinical trials—FEM-PrEP and VOICE were unable to show protection from HIV infection in heterosexual women [24, 25]. Possible biological explanations for this show tenofovir levels ten-fold higher in rectal tissue than in vaginal tissue [26], indicating decreased protection by PrEP in women. These divergent outcomes call for further investigation of the disposition of tenofovir and emtricitabine to ensure complete efficacy of PrEP in all populations.

III. Nucleotide reverse transcriptase inhibitors by are activated by nucleotide kinases

Tenofovir requires two phosphorylation steps to form the competitive HIV reverse transcriptase inhibitor and nucleotide triphosphate analog, tenofovir diphosphate [17]. Tenofovir is first phosphorylated intracellularly to tenofovir-monophosphate [17]. Tenofovir-monophosphate then is phosphorylated to tenofovir-diphosphate (TFV-DP), which is the pharmacologically active form of TFV. Previous work done in our lab

interrogated this phosphorylation pathway, demonstrating not only the kinases involved in tenofovir activation, but also the compartment specificity of these kinases [27].

Phosphorylation of tenofovir to tenofovir monophosphate has been observed from adenylate kinase 2 *in vitro* using purified protein from mouse and human cell lines [17, 28]. *In vivo*, we have demonstrated phosphorylation of tenofovir by adenylate kinase 2 in peripheral blood mononuclear cells, vaginal, and colorectal tissues (Figure 1) [27]. Further, our lab has shown phosphorylation of tenofovir-monophosphate to tenofovir-diphophate by creatine kinase, muscle in colorectal tissue [27]. Pyruvate kinase, muscle and pyrivate kinase, liver and red blood cell catalyze the phosphorylation of tenofovir-monophosphate to tenofovir-diphosphate in peripheral blood mononuclear cells and vaginal tissue [27].

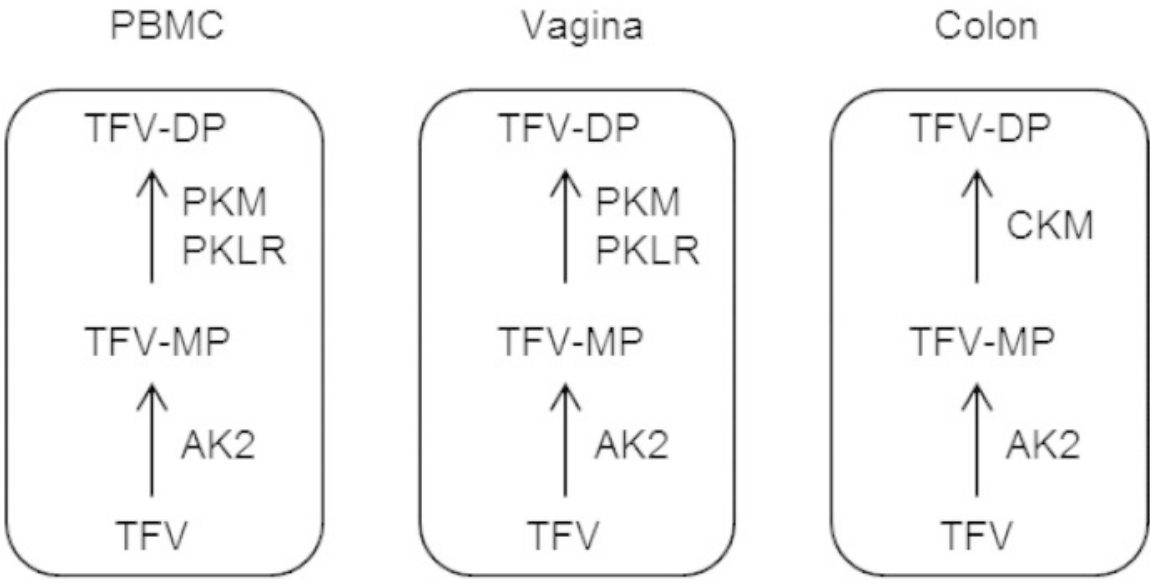

Figure 1. Proposed sequence of tenofovir phosphorylation to form the active nucleotide triphosphate analog competitive inhibitor of HIV reverse transcriptase, taken from Lade et al. [27]

Emtricitabine, like tenofovir, is also nucleoside/nucleotide reverse transcriptase inhibitor that requires intracellular phosphorylation. In contrast to the monophosphate analog tenofovir, however, which needs to be phosphorylated twice, emtricitabine requires three phosphorylation steps to form nucleotide triphosphate analog necessary to inhibit HIV reverse transcriptase [18]. We thus put forward four candidate kinases that could activate emtricitabine in HIV susceptible tissues. Deoxycytidine kinase purified from calf thymus has demonstrated *in vitro* phosphorylation of emtricitabine to emtricitabine-monophosphate [18]. Thymidine kinase 1, another candidate kinase for emtricitabine phosphorylation has shown activity toward zidovudine and stavudine, nucleoside analogs that are structurally similar to emtricitabine [29]. Unlike emtricitabine, a cytidine analog, zidovudine and stavudine are thymidine analogs. However, all compounds are dideoxynucleosides, lacking both 2'- and 3'-hydroxyl groups in their sugar ring. Emtricitabine-monophosphate is proposed to be phosphorylated to emtricitabine-diphosphate from activity of the cytidine monophosphate kinase 1 toward lamivudine-monophosphate [30]. Lamivudine, like emtricitabine, is a cytidine analog. Differences between the structures of the two compounds are scarce, with the only difference being a fluorine atom in emtricitabine at the 5 position of the cytidine base. Phosphorylation of lamivudine-monophosphate to lamivudine diphosphate using purified human cytidine monophosphate kinase 1 has been demonstrated, indicating possible activity toward emtricitabine-monophosphate [30]. To form the pharmacologically active metabolite emtricitabine-triphosphate, we predict phosphoglycerate kinase 1 to demonstrate activity toward emtricitabine-diphosphate. This hypothesis stems from previously demonstrated phosphorylation of the anabolite of

the deoxynucleoside analog L-Fd4C [31]. Like lamivudine, this compound shares structural similarities with emtricitabine. Differing between L-Fd4C and emtricitabine is a double bond between the pentose 2' and 3' positions rather than a sulfur atom at the 3' position found on emtricitabine. To demonstrate activity of phosphoglycerate kinase 1 toward L-Fd4C, this kinase was purified from HepG2 cells [31].

IV. Pharmacogenetics may contribute to differences observed in PrEP outcomes

Despite taking into account participant adherence and demonstrating PrEP efficacy, prior PrEP clinical trials are unable to establish complete protection from HIV infection. Additional investigation of potential factors contributing to the divergent results of these clinical trials is necessary. One possible source of this divergence is the role of genetic variants in drug activation. Genetic variation in drug-metabolizing enzymes has been shown to play a role in inter-individual metabolism of and response to antiretrovirals [32].

Inter-individual genetic variation has been demonstrated in part in studies examining the pharmacokinetics of efavirenz, a non-nucleoside reverse transcriptase inhibitor [33, 34]. In a study comparing drug exposure in a black, non-Hispanic population and a European-American population, efavirenz was found to have a longer elimination half-life [33]. Geographic differences in efavirenz metabolism were also observed in another study, with individuals in Thailand and South Africa exhibiting lower clearance of efavirenz than in individuals in Europe or America [34].

Studies carried out in our laboratory examining the anti-HIV drug maraviroc have demonstrated dissimilarities in drug metabolism among individuals exhibiting differing CYP3A5 genotypes [35]. In this pharmacokinetic study, maraviroc exposure was found to be reduced in individuals that were homozygous for the CYP3A5*1 allele [35]. This lower exposure, in comparison to individuals homozygous for the CYP3A5 wild-type allele, indicates maraviroc may be unsuitable for treatment against HIV in individuals homozygous for the CYP3A5*1 allele.

Unlike the efavirenz and maraviroc, tenofovir is not metabolized by any cytochrome P450 enzymes [36]. However, tenofovir must be activated intracellularly by kinases to form the nucleotide triphosphate analog that will competitively inhibit HIV reverse transcriptase. Given the evidence of inter-individual drug exposure previously discussed, one can envision the presence of genetic variants in kinases that activate tenofovir to play a role in tenofovir activation. A prior study in our lab has found 71 previously unreported genetic variants in the genes that encode the kinases that have been found to phosphorylate tenofovir to tenofovir diphosphate [27]. These detected variants could negatively impact tenofovir activation, ultimately compromising the efficacy of HIV pre-exposure prophylaxis despite complete participant adherence.

V. MALDI-MS Imaging informs spatial distribution of drug disposition

The compartment specificity of tenofovir activation demonstrated in our previous work [27] and investigations of the efficacy of topical PrEP formulations [26, 37] require

a deeper understanding of tenofovir tissue pharmacokinetics. Compartments at risk for HIV infection such as rectal, vaginal, and cervical tissues [26, 37, 38, 39] have been interrogated for their concentration of tenofovir and its active anabolite tenofovir disphosphate. Using liquid chromatography coupled to tandem mass spectrometry, these studies have been helpful in confirming and measuring the amount of drug able to penetrate HIV-susceptible tissues, but are not able to provide any distribution or localization information. Because spatial distribution information is lost using this method, these studies have only been able to provide average quantifications of tenofovir and tenofovir diphosphate across the tissue sample. This method of quantification requires homogenization of the collected tissue sample, destroying any spatial distribution information available across tissue. In contrast to liquid chromatography-tandem mass spectrometry techniques, spatial distribution of a drug can be preserved when using matrix assisted laser desorption/ionization coupled to imaging mass spectrometry (MALDI-IMS) [40]. Rather than necessitating the use of a homogenate, tissue can be analyzed in MALDI-IMS by way of tissue slices mounted on a glass slide. Because homogenization of tissue is not necessary for analysis using this method, MALDI-IMS can therefore be applied to describe drug distribution in tissue.

References

[1] UNAIDS, "UNAIDS Global AIDS Update 2016.," UNAIDS, 2016.

[2] S. D.-B. K. Weller, "Condom effectiveness in reducing heterosexual HIV transmission.," *Cochrane Database of Systematic Reviews,* 2002.

[3] M. S. P. C. e. a. Bailey RC, "Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial.," *Lancet,* vol. 369, no. 9562, pp. 643-656, 2007.

[4] K. G. S. D. e. a. Gray RH, "Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial.," *Lancet,* vol. 369, no. 9562, pp. 657-666, 2007.

[5] T. D. L. E. e. a. Auvert B, "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial.," *PLoS Med.,* vol. 2, no. 11, 2005.

[6] H. M. J. D. e. a. Chasela CS, "Maternal or infant antiretroviral drugs to reduce HIV-1 transmission," *New England Journal of Medicine,* vol. 362, pp. 2271-2281, 2010.

[7] M. T. Y. N. Nakasone T, "Double oral administration of emtricitabine/tenofovie prior to virus exposure protects against highly pathogenic simian/human immunodeficiency virus infection in macaques.," *Jpn J Infect Dis,* vol. 65, pp. 345-349, 2012.

[8] G. RH., "CDC Releases Updated Guidelines for Postexposure Prophylaxis After Sexual, Injection Drug, or Other Nonoccupational Exposures to HIV," *American Family Physician,* vol. 94, no. 5, pp. 392-393, 2016.

[9] C. Y. M. M. G. T. H. M. K. N. H. J. K. J. G. B. P. J. G. S. M. S. C. S. S. B. M. K. H. I. E. S. P.-M. E. W. L. M. J. M. L. Cohen MS, "Prevention of HIV-1 infection with early antiretroviral therapy.," *The New England Journal of Medicine,* vol. 365, no. 6, pp. 493-505, 2011.

[10] M. K. S. M. P. K. K. A. Cohen MS, "Antiviral agents and HIV prevention: controversies, conflicts, and consensus.," *AIDS,* vol. 26, no. 13, pp. 1585-1598, 2012.

[11] O. R. Q. S. e. a. Garcia-Lerma JG, "Prevention of Rectal SHIV Transmission in Macaques by Daily or Intermittent Prophylaxis with Emtricitabine and Tenofovir," *PLoS Med.,* vol. 5, no. 2, 2008.

[12] D. C. S. S. e. a. Parikh UM, "Complete protection from repeated vaginal simian-human immunodeficiency virus exposures in macaques by a topical gel containing tenofovir alone or with emtricitabine.," *J. Virol.,* vol. 83, pp. 10358-10365, 2009.

[13] Y. A. Z. Q. e. a. Cong ME, "Protection against rectal transmission of an emtricitabine-resistant simian/human immunodeficiency virus SHIV162p3M184V mutant by intermittent prophylaxis with Truvada.," *J. Virol.,* vol. 85, pp. 7933-7936, 2011.

[14] C. C. Baeten J, "Systemic and topical drugs for the prevention of HIV infection: antiretroviral pre-exposure prophylaxis.," *Annu. Rev. Med.,* vol. 64, pp. 219-232, 2013.

[15] CDER-USFDA., "Approval Package for: 021752Orig1s030.," July 2012. [Online]. Available: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/021752Orig1s030Approv.pdf. [Accessed October 2016].

[16] U. F. a. D. Administration, "FDA Approves First Medication to Reduce HIV Risk.," FDA Consumer Health Information, 2012.

[17] B. S. R. K. C. B. N. F. A. Robbins, "Anti-human immunodeficiency virus activity and cellular metabolism of a potential prodrug of the acyclic nucleoside phosphonate 9-R-(2-phosphonomethoxypropyl)adenine (PMPA).," *Antimicrobial Agents and Chemotherapy,* vol. 42, no. 3, pp. 612-617, 1998.

[18] P. D. M. L. D. P. M. e. a. Furman, "The Anti-Hepatitis B Virus Activities, Cytotoxicities, and Anabolic Profiles of the (-) and (+) Enantiomers of cis-5-Fluoro-1-[2-(Hydroxymethyl)-1 ,3-Oxathiolan-5-yl]Cytosine.," *Antimicrobial Agents and Chemotherapy.,* vol. 36, no. 12, pp. 2696-2692, 1992.

[19] J. Baeten, D. Donnell, P. Ndase and t. P. P. S. Team, "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women.," *N. Engl. J. Med,* vol. 367, pp. 399-410, 2012.

[20] R. Grant, L. KR, A. PL and e. al., "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men.," *N. Engl. J. Med.,* vol. 363, no. 27, pp. 2587-2599, 2010.

[21] M. Thigpen, K. PM, P. LA, t. T. S. Group and e. al., "Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana.," *N. Engl. J. Med.,* vol. 367, pp. 423-434, 2012.

[22] V. S. S. P. e. a. Martin M, "Enrollment characteristics and risk behaviors of injection drug users participating in the Bangkok Tenofovir Study, Thailand.," *PLoS ONE.,* vol. 6, 2011.

[23] M. M. S. P. S. U. M. P. L. M. C. S. K. P. N. P. K. S. C. R. G. R. M. J. P. L. C. M. H. C. V. S. Choopanya K and B. T. S. Gr, "Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial.," *Lancet,* vol. 381, no. 9883, pp. 2083-2090, 2013.

[24] C. A. A. K. A. K. L. J. K. S. M. M. O. F. M. R. O. J. T. L. M. M. M. P. M. M. R. I. M. S. S. L. K. H. K. S. W. C. G. R. K. A. N. K. M. J. Van Damme L, "Preexposure prophylaxis for HIV infection among African women.," *N Engl J Med.,* vol. 367, no. 5, pp. 411-422, 2012.

[25] J. e. a. Marrazzo, "Pre-exposure prophylaxis for HIV in women: daily oral tenofovir, oral tenofovir/emtricitabine, or vaginal tenofovir gel in the VOICE study (MTN 003)."," in *0th Conference on Retroviruses and Opportunistic infections.* , Atlanta, GA., 2013.

[26] K. P. H. K. E. J. A. e. a. Patterson, "Penetration of Tenofovir and Emtricitabine in Mucosal Tissues: Implications for Prevention of HIV-1 Transmission," *Science Translational Medicine,* vol. 3, no. 112, 2011.

[27] T. E. H. C. ,. B. N. Lade JM, "Discovery of Genetic Variants of the Kinases That Activate Tenofovir in a Compartment-specific Manner," *EBioMedicine,* vol. 2, no. 9, p. 1145–1152, 2015.

[28] H. K. V. I. H. A. Krejcova R, "Phosphorylation of purine (phophonomethoxy)alkyl derivatives by mitochondrial AMP kinase (AK2 type) from L1210 cells.," *Collection of Czechoslovak Chemical Contributions. ,* vol. 65, pp. 1653-1668, 2000.

[29] H. P. D. C. E. Balzarini J, "Differential patterns of intracellular metabolism of 2',3'-didehydro-2',3'-dideoxythymidine and 3'-azido-2',3'-dideoxythymidine, two potent anti-human immunodeficiency virus compounds.," *J Biol Chem. ,* vol. 264, no. 11, pp. 6127-6133, 1989.

[30] P. e. al., "Reaction of human UMP-CMP kinase with natural and analog substrates," *Eur. J. Biochem. ,* vol. 270, p. 1784–1790, 2003.

[31] P. F. Q. e. a. Krishnan, "Phosphorylation of Pyrimidine Deoxynucleoside Analog Diphosphates: SELECTIVE PHOSPHORYLATION OF L-NUCLEOSIDE ANALOG DIPHOSPHATES BY 3-PHOSPHOGLYCERATE Kinas.," *The Journal of Biological Chemistry. ,* vol. 277, no. 7, pp. 5453-5459, 2002.

[32] B.-M. T. T. J. F. D. D. Z. W. M. Michaud V, "The dual role of pharmacogenetics in HIV treatment: mutations and polymorphisms regulating antiretroviral drug resistance and disposition.," *Pharmacol Rev. ,* vol. 64, no. 3, pp. 803-833, 2012.

[33] L. L. H. S. M. J. B. K. R. S. S. L. Pfister M, " Population pharmacokinetics and pharmacodynamics of efavirenz, nelfinavir, and indinavir: adult AIDS Clinical Trial Group Study 398.," *Antimicrob Agents Chemother,* vol. 47, pp. 130-137, 2003.

[34] v. L. F. M. T. L. J. B. J. H. A. Kappelhoff BS, "Nevirapine and efavirenz pharmacokinetics and covariate analysis in the 2NN study.," *Antivir Ther,* vol. 10, pp. 145-155, 2005.

[35] F. E. H. C. B. N. Lu Y, "CYP3A5 genotype impacts maraviroc concentrations in healthy volunteers.," *Drug Metab Dispos.,* vol. 42, no. 11, pp. 1796-1802, 2014.

[36] J. Baeten, "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women.," *N Engl J Med. ,* vol. 367, no. 5, pp. 399-410. , 2012.

[37] M. C. R. Reyzer, "MALDI-MS-based imaging of small molecules and proteins in tissues.," *Curr Opin Chem Biol. ,* vol. 11, no. 1, pp. 29-35, 2007.

[38] V. K. H. B. R. V. G. S. S. H. Kerstin Strupat, "MALDI Produced Ions Inspected with a Linear Ion Trap-Orbitrap Hybrid Mass Analyzer," *Journal of the American Society for Mass Spectrometry,* vol. 20, no. 8, pp. 1451-1463, 2009.

Chapter 2: Discovery of genetic variants of the kinases that activate tenofovir among individuals from the United States, Thailand and South Africa

Abstract

Tenofovir (TFV) is a nucleotide reverse transcriptase inhibitor approved for use in HIV pre-exposure prophylaxis (PrEP), a pharmacological approach to HIV prevention. TFV requires two phosphorylation steps to form a competitive inhibitor of HIV reverse transcriptase. The kinases adenylate kinase 2 (*AK2*), creatine kinase, muscle (*CKM*), pyruvate kinase, muscle (*PKM*) and pyruvate kinase, liver and red blood cell (*PKLR*) have been identified to phosphorylate tenofovir in peripheral blood mononuclear cells (PBMC), vaginal, and colorectal tissue. This study employed next-generation sequencing of genomic DNA isolated from 505 clinical trial participants in order to probe for variants of the genes encoding these kinases. Using bioinformatics tools, genetic variants predicted to impact TFV activation were identified in subjects across three geographic sites, namely Bangkok, Thailand, Cape Town, South Africa and New York City, USA. We identified 103 novel unreported variants across the kinases *AK2, CKM, PKM* and *PKLR,* with 19 of 505 individuals (3.7% frequency) carrying variants that were predicted to be deleterious and probably damaging. Investigating overlap of variants across the three geographic locations where sequencing for variants in *AK2, CKM, PKM,* and *PKLR* was carried out, 2 out of 19 *AK2* variants were shared among individuals located in Bangkok and Cape Town, and in Bangkok and New York City. For *PKLR*, 1 out of 41 variants were detected in both Cape Town and New York City. These data suggest that genetic variants exist in these populations that could potentially impact TFV

phosphorylation, contributing to variable outcomes in PrEP trials that cannot be explained by adherence alone.

Introduction

Tenofovir (TFV) is a nucleotide reverse transcriptase inhibitor currently used for the treatment and prevention of human immunodeficiency virus (HIV) infection [1]. In fact, Truvada ®, which is a combination of TFV and the nucleoside reverse transcriptase inhibitor emtricitabine is currently the only drug approved by the FDA for use in HIV pre-exposure prophylaxis (PrEP) [2]. PrEP is a pharmacological strategy for preventing HIV infection, the efficacy of which has been bolstered by 3 large-scale clinical trials including the iPrEX study comprised of 2,499 men who have sex with men (MSM) or transgender women (TGW) [3], the Botswana TDF2 study comprised of 1,219 men and women [4], and the Partners PrEP study, comprised of 4,747 serodiscordant couples [5]. Through examining significant populations at-risk for HIV infection, these trials demonstrated protection against HIV through daily oral administration of Truvada ®. When adjusted for adherence, 92% of MSM or TGW in the iPrEx study were observed to be protected from HIV infection in comparison to the placebo group [3]. The Botswana TDF2 study HIV-uninfected men and women, 78% of whom were protected against HIV when adjusted for adherence [4]. Finally, in the largest of these studies, Partners PrEP, a 90% decrease in HIV infection among HIV-uninfected partners in comparison to placebo was observed [5]. However, despite taking adherence into account in the aforementioned studies, full protection against HIV was not demonstrated.

In order to produce the active HIV reverse trancriptase inhibitor, TFV must be phosphorylated twice, first to tenofovir-monophosphate (TFV-MP) [6]. TFV-MP then is phosphorylated to tenofovir-diphosphate (TFV-DP), which is the pharmacologically active form of TFV. Our lab has investigated this phosphorylation pathway and demonstrated the compartment specificity of the kinases involved in TFV activation [7]. Specifically, we found that adenylate kinase 2 (AK2) phosphorylates TFV to TFV-MP in peripheral blood mononuclear cells (PBMC), vaginal, and colorectal tissue. Interestingly, while we observed phosphorylation of TFV-MP to TFV-DP by creatine kinase, muscle (CKM) in colorectal tissue, the kinases pyruvate kinase, muscle (PKM) and pyruvate kinase, liver and red blood cell (PKLR) catalyzed this step in PBMC and vaginal tissue. Through this work, we proposed the possibility that genetic variants of these kinases could impact the kinase function and therefore TFV phosphorylation. In doing so, via genotyping 142 clinical trial participants across geographic regions located in the United States, South Africa and Uganda, we identified 71 previously unreported genetic variants in the genes encoding *AK2, CKM, PKM,* and *PKLR*. In the present study, we extend this work substantially through the analysis of genomic DNA isolated from 505 clinical trial participants across three geographic locations, namely Bangkok, Thailand (n = 171), New York City, USA (n = 149), and Cape Town, South Africa (n = 185). Of particular note, this study represents the first targeted resequencing analysis of the *AK2, CKM, PKM,* and *PKLR* genes in subjects in Asia. This work further establishes that genetic deficiencies in the ability to activate TFV may exist in certain individuals.

Materials and Methods

Clinical study sites and sample collection. Dried blood spots (DBS) were obtained from HIV-uninfected individuals (n = 505) enrolled in the HIV Prevention Trials Network study HPTN 067 across three clinical research sites (CRS): Emavudleni CRS in Cape Town, South Africa; Silom Community Clinic CRS in Bangkok, Thailand, and the Harlem Prevention Center CRS (El-Sard CTU) in New York, United States. All participants sequenced consented to genetic testing. Every individual included in this study contributed five DBS, with each DBS containing 50 μL whole blood spotted onto Whatman™ 903 Protein Saver Cards. Information about participants enrolled and genotyped in this study is summarized in Table 1. These participants were MSM or TGW and women who have sex with men (WSM) from three geographic locations.

HPTN 067 study locations and participant demographics n = 505		
Study site	Reported gender	n (% of n)
Bangkok, Thailand	MSM, TGW	171 (34)
New York City, USA	MSM, TGW	149 (30)
Cape Town, South Africa	WSM	185 (36)

Table 1. HPTN 067 participant enrollment location and gender information.

Study participants were enrolled across three geographic locations, with MSM and TGW enrolled in Bangkok (n = 171) and New York City (n = 149) and WSM enrolled in Cape Town (n = 185).

Genomic DNA isolation from HPTN 067 DBS samples. Genomic DNA was isolated from 2 DBS from each individual using a QIAamp 96 DNA Blood Kit (QIAGEN, Valencia, CA). DBS were punched out using a hole puncher. Genomic DNA was extracted following the supplementary protocol "Isolation of genomic DNA from dried blood spots using the QIAamp ®96 DNA Blood Kit – (EN)" from QIAGEN. Purified DNA was eluted using 150 µL DEPC-treated, nuclease-free water (Quality Biological, Inc., Gaithersburg, MD) and concentrated using a ZR-96 DNA Clean-up Kit™ (Zymo Research, Irvine, CA). Resulting concentrated and purified genomic DNA was eluted in 12 µL DEPC-treated, nuclease-free water.

Next-generation sequencing target design, sample preparation, and analysis. Sequencing of the kinases *AK2*, *CKM*, *PKM* and *PKLR* was executed using the Illumina TruSeq Custom Amplicon kit v1.5 (Illumina, San Diego, CA). Probes for sequencing the combination of these kinases were designed using Illumina DesignStudio software as previously described [7].Genomic DNA isolated from clinical samples was processed

following the Illumina TruSeq Custom Amplicon Library Preparation Guide (Part Number 15027983 Rev. C, August 2013). DNA concentration was measured using a Qubit ® 3.0 Fluorimeter (Thermo Scientific, New York, NY). Fifty ng of DNA input were used per DNA sample sequenced. The resulting prepared DNA library (6 µL) was diluted in 594 µL HT1 buffer containing 1% PhiX sequencing control. Illumina VariantStudio software was used to annotate and analyze variant read quality as previously described [7].

Results

Geographic locations of clinical trial participants investigated for the presence of genetic variants. In order to probe the existence of variants in the genes that encode the kinases that activate TFV, DBS from 505 participants in Bangkok, Thailand, Cape Town, South Africa, and New York City, USA were investigated. Of the 505 participants sequenced for *AK2, CKM, PKM,* and *PKLR* in this study, 171 were located in Bangkok, Thailand, 149 were located in New York City, USA, and 185 were located in Cape Town, South Africa.

Analysis of individuals carrying genetic variants in tenofovir-activating kinases *AK2, CKM, PKM,* and *PKLR*. To examine the presence of genetic variants in the nucleotide kinases reported to activate TFV, we employed a previously described targeted assay to sequence the exonic regions of the genes *AK2, CKM, PKM* and *PKLR* [7]. Further, we sequenced participants in each of the three geographic locations Bangkok (n = 171), New York City (n = 149), and Cape Town (n = 185) included in the HPTN 067 clinical study. Variant distribution among Bangkok study participants is

demonstrated using a Venn diagram in Fig. 1. At the Bangkok study site, 29 of the 171 participants sequenced (17%, 29/171) for these kinases carried single nucleotide variants (SNVs) or deletions that were predicted to result in a mutation at the amino acid level. Of these 29 participants from the Bangkok study site, 26 individuals exhibited variants for a singular nucleotide kinase. Moreover, one individual had a variant detectable for only *AK2*, nine individuals had variants detectable for only *CKM*, eight individuals had variants detectable for only *PKM,* and eight individuals had variants detectable for only *PKLR.* Of the remaining three participants, one individual carried variants detectable in *AK2* and *PKM* and two individuals had genetic variants in both *PKM* and *PKLR*.

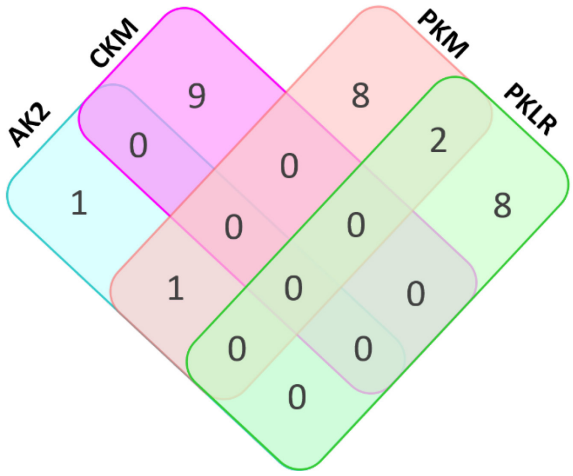

Figure 1. Distribution of individuals enrolled at the Bangkok study site carrying genetic variants in TFV-activating kinases. Each rectangle is representative of a TFV-activating kinase that was sequenced: *AK2* in blue, *CKM* in pink, *PKM* in orange and *PKLR* in green. Numerical values indicate the number of individuals detected to carry a single nucleotide variation or deletion. Overlapping regions of each rectangle indicate the number of individuals with genetic variants in more than one kinase.

At the New York City study site, 35 of the 149 participants sequenced (23%, 35/149) were observed to carry SNVs or deletions in predicted to result in a mutation at the amino acid level. The observed variant distribution among New York City study

participants is depicted using a Venn diagram in Fig. 2. Of the 35 participants from the New York City study site detected to carry a genetic variant, 27 exhibited variants for only one of the kinases sequenced. Six individuals carried variants for *AK2* alone, seven individuals for *CKM* alone, six individuals for *PKM* alone, and eight individuals carried variants for *PKLR* alone. Further, of these 35 participants, eight carried genetic variants in more than one kinase. One individual was observed to have detectable variants in *AK2, PKM,* and *PKLR,* two individuals were observed to have detectable variants in *AK2* and *PKLR*, two individuals were observed to have detectable variants in *CKM* and *PKM* and three individuals were observed to carry variants in *PKM* and *PKLR*.

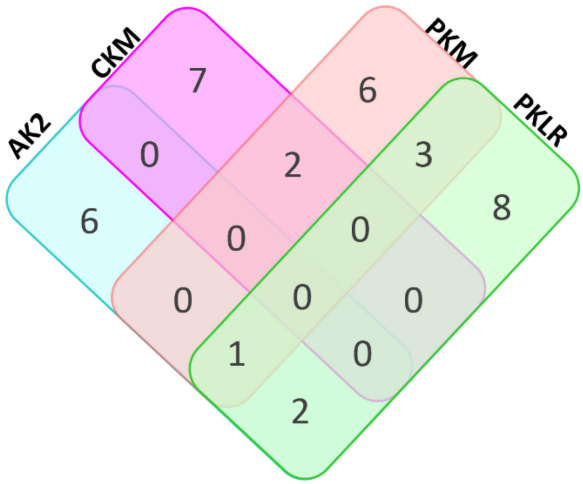

Figure 2. Distribution of individuals enrolled at the New York City study site carrying genetic variants in TFV-activating kinases. Each rectangle represents a TFV-activating kinase that was sequenced: *AK2* in blue, *CKM* in pink, *PKM* in orange and *PKLR* in green. Numerical values in each compartment represent the number of individuals detected to have a single nucleotide variation or deletion. Overlapping regions of each rectangle indicate the number of individuals with genetic variants in more than one kinase.

Variant distribution among Cape Town study participants is shown using a Venn diagram in Fig. 3. Thirty of the 185 participants (16%, 30/185) enrolled at this study site

were observed to have SNVs or deletions predicted to result in a mutation at the amino acid level. Twenty-two individuals of the 30 individuals carrying variants in Cape Town exhibited single nucleotide variants for only one of the kinases sequenced. Of these 22 individuals, two displayed variants in the kinase *AK2* alone, seven in the kinase *CKM* alone, five in the kinase *PKM* alone, and eight in the kinase *PKLR* alone. In contrast to exhibiting variants in only one kinase, eight participants carried genetic variants in more than one kinase. Two individuals were observed to have detectable variants in *AK2* and *PKM*. One individual was observed to have detectable variants in both *AK2* and *CKM*. Seven individuals were observed to have detectable variants in only *CKM,* while two individuals were observed to have detectable variants in *CKM* and *PKM* and two individuals were observed to have detectable variants in *CKM* and *PKLR*. Five individuals were observed to have detectable variants in *PKM*. Eight individuals were observed to have detectable variants in *PKLR*. One individual was observed to carry detectable variants in *PKM* and *PKLR*.

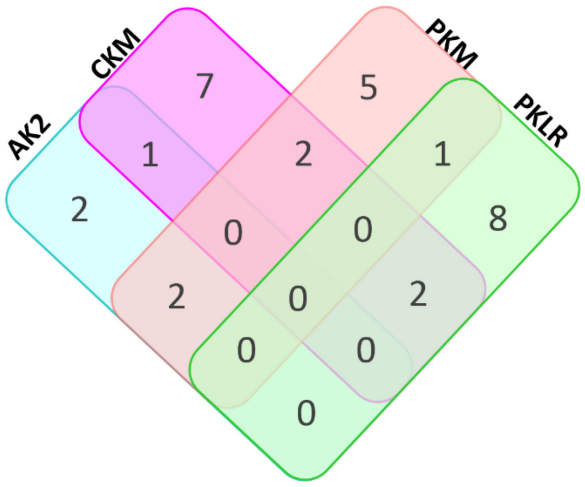

Figure 3. Distribution of individuals enrolled at the Cape Town study site with detectable genetic variants in TFV-activating kinases. Each rectangle represents a TFV-activating

kinase that was sequenced: *AK2* in blue, *CKM* in pink, *PKM* in orange and *PKLR* in green. Numerical values in each compartment represent the number of individuals detected to have a single nucleotide variation or deletion. Overlapping regions of each rectangle indicate the number of individuals with genetic variants in more than one kinase.

Investigation of *AK2* genetic variants detected across three geographic locations. Our laboratory has previously demonstrated *AK2* phosphorylation of TFV to TFV-MP in PBMC, vaginal tissue, and colon tissue. In the 505 participants sequenced in this study, we observed 17 single nucleotide variations (SNVs) predicted to result in a mutation of the amino acid sequence for the DNA reference sequence corresponding to *AK2* (NM_001625.3). These specific variants were detected across 16 individuals. The described variants, which include both previously reported and novel variants, are detailed in Supplementary Table 1. All 16 individuals were observed to be heterozygous for the variants detected. Further, of the 16 individuals observed to carry at least one SNV, two individuals were of the Bangkok study site, five individuals were of the Cape Town study site, and nine individuals were of the New York City study site. Among the individuals carrying at least one SNV, one individual at the Cape Town study site was observed to be carrying two variants while two individuals at the New York City study site were observed to be carrying two variants. Of the 17 *AK2* genetic variants observed, 13 of these variants, which were observed across 11 individuals, were previously unreported (Table 2). Of the 505 individuals sequenced for *AK2* missense variants, one individual in New York City was observed to have the previously unreported variant at the coding DNA position c.664G>A. Applying *in silico* functional prediction tools to the variant c.664G>A, it is predicted to be “deleterious” and “probably damaging”. *In silico* tools SIFT and PolyPhen were used to predict the functional consequences of these

variants as previously described [7]. Four variants contained in the Single Nucleotide Polymorphism database (dbSNP), were detected and are listed as follows: rs143825456 (NM_001625.3:c.631G>A) was detected in two individuals, with one individual located in Bangkok and the other in Cape Town; rs148421308 (NM_001625.3:c.460G>A) was detected in one individual in New York City; rs61750965 (NM_001625.3:c.386G>A) was detected in one individual in New York City; rs12116440 (NM_001625.3:c.625G>A) was detected in one individual in New York City.

Geographic Location	Variant (ref.>alt.)	cDNA Position	Coding DNA Sequence Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon	SIFT Prediction	PolyPhen Prediction
Bangkok	T>C	699	616	206	I>V	6/6	tolerated(0.38)	benign(0.063)
Cape Town	G>A	106	23	8	A>V	1/6	tolerated(0.27)	benign(0.002)
Cape Town	T>C	166	83	28	K>R	1/6	deleterious(0)	possibly_damaging(0.693)
Cape Town	G>A	247	164	55	A>V	2/6	deleterious(0)	possibly_damaging(0.774)
Cape Town	C>T	258	175	59	E>K	2/6	tolerated(0.29)	benign(0.01)
Cape Town	C>T	312	229	77	E>K	3/6	deleterious(0.01)	possibly_damaging(0.656)
New York City	A>C	139	56	19	V>G	1/6	deleterious(0)	possibly_damaging(0.882)
New York City	C>T	237	154	52	A>T	2/6	deleterious(0.01)	possibly_damaging(0.859)
New York City	T>C	267	184	62	K>E	2/6	tolerated(0.59)	benign(0.134)
New York City	C>T	546	463	155	E>K	5/6	tolerated(0.59)	benign(0.187)
New York City	G>A	664	581	194	T>I	6/6	deleterious(0)	probably_damaging(0.997)
New York City	T>C	699	616	206	I>V	6/6	tolerated(0.38)	benign(0.063)
New York City	T>A	699	616	206	I>F	6/6	deleterious(0.03)	benign(0.288)

Table 2. Previously unreported *AK2* missense variants detected in clinical trial participants located in Bangkok, Thailand, Cape Town, South Africa, and New York City, USA.

Thirteen previously unreported genetic missense variants were detected in 11 individuals out of 505 individuals sequenced for the coding DNA reference sequence NM_001625.3. All individuals that had detectable variants were heterozygous for those variants. One individual from Cape Town had two detectable missense variants and one individual from New York City had two detectable missense variants. Across the three geographic locations where individuals were sequenced for *AK2*, one *AK2* variant was found in both Bangkok and New York City but not in Cape Town.

Investigation of sequence variants of *CKM* detected across three geographic locations. Because *CKM* has been shown to phosphorylate TFV-MP to TFV-DP in colon tissue, the existence of genetic variants in *CKM* predicted to yield functional effects on this kinase was investigated using a targeted next-generation sequencing assay. For the DNA reference sequence corresponding to *CKM* (NM_001824.4), 34 total variants comprised of SNVs and deletions that are predicted to be reflected in the amino acid sequence were detected across 30 individuals. The total observed variants, including previously reported and previously unreported variants, are listed in Supplementary Table

1. Twenty-nine individuals were observed to be heterozygous for all the variants detected, while one individual in New York City was observed to be homozygous. Of the 30 individuals observed to have detectable SNVs or deletions, nine individuals were located at the Bangkok study site, 12 individuals were located at the Cape Town study site, and nine individuals were located at the New York City study site. We observed two clinical trial participants exhibiting more than one variant for this kinase. One individual at the Bangkok study site was observed to carry two different variants and one individual at the Cape Town study site was observed to carry two variants. Twenty-six previously unreported single nucleotide variants or base deletions were detected across 24 individuals (Table 3). Of the 505 individuals sequenced for *CKM* missense variants, nine individuals were observed to have a predicted "deleterious" and "probably damaging" variant, yielding a frequency of 1.8%. *In silico* tools SIFT and PolyPhen were used to predict the functional consequences of these variants. Five reported variants, found in the dbSNP, were detected and are listed as follows: rs149354459 (NM_001824.4:c.752G>A) was detected in one individual from Bangkok, rs17875625 (NM_001824.4:c.728G>T) was detected in one individual from Bangkok; rs201048164 (NM_001824.4:c.163G>A) was detected in one individual from Cape Town; rs17850202 (NM_001824.4:c.673T>C) was detected in one individual from New York City and rs11559024 (NM_001824.4:c.248A>G) was detected in one individual from New York City.

Geographic Location	Variant (ref.>alt.)	cDNA Position	Coding DNA Sequence Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon	SIFT Prediction	PolyPhen Prediction
Bangkok	C>T	303	128	43	R>Q	2/8	tolerated(0.07)	benign(0.288)
Bangkok	T>C	693	518	173	Y>C	5/8	deleterious(0)	probably_damaging(0.991)
Bangkok	T>A	696	521	174	Y>F	5/8	tolerated(0.18)	possibly_damaging(0.717)
Bangkok	T> C	713	538	180	T>A	5/8	deleterious(0)	benign(0.086)
Bangkok	C> A	977	802	268	G>C	7/8	deleterious(0)	possibly_damaging(0.822)
Bangkok	T> G	1262	1087	363	M>L	8/8	tolerated(0.35)	benign(0.08)
Cape Town	G> C	451	276	92	I>M	3/8	deleterious(0)	probably_damaging(0.951)
Cape Town	G> A	569	394	132	R>C	4/8	deleterious(0)	probably_damaging(0.999)
Cape Town	C> T	623	448	150	E>K	4/8	deleterious(0.01)	possibly_damaging(0.586)
Cape Town	C> T	638	463	155	E>K	4/8	deleterious(0)	benign(0.177)
Cape Town	T> C	968	793	265	K>E	7/8	deleterious(0.04)	possibly_damaging(0.568)
Cape Town	A> G	990	815	272	M>T	7/8	deleterious(0.05)	possibly_damaging(0.503)
Cape Town	C> A	991	816	272	M>I	7/8	tolerated(0.05)	benign(0.142)
Cape Town	T> A	995	820	274	N>Y	7/8	deleterious(0)	probably_damaging(0.91)
Cape Town	C> A	1013	838	280	V>L	7/8	tolerated(0.18)	benign(0.159)
Cape Town	T> A	1032	857	286	N>I	7/8	deleterious(0)	probably_damaging(1)
Cape Town	C> G	1064	889	297	V>L	7/8	tolerated(0.21)	benign(0.299)
Cape Town	G> T	1124	949	317	L>M	7/8	deleterious(0.02)	probably_damaging(0.998)
New York City	C> A	392	217	73	G>C	3/8	deleterious(0)	probably_damaging(0.997)
New York City	T> A	435	260	87	E>V	3/8	deleterious(0.02)	benign(0.402)
New York City	T> A	482	307	103	T>S	3/8	tolerated(0.62)	benign(0.001)
New York City	T> C	720	545	182	K>R	5/8	tolerated(0.11)	benign(0.017)
New York City	A> T	806	631	211	W>R	5/8	deleterious(0)	probably_damaging(0.994)
New York City	T> G	837	662	221	D>A	6/8	tolerated(0.08)	benign(0.148)
New York City	A> G	924	749	250	F>S	6/8	deleterious(0)	probably_damaging(1)
New York City	C> T	1172	997	333	V>I	8/8	tolerated(0.3)	benign(0.085)

Table 3. Previously unreported *CKM* missense variants detected in clinical trial participants located in Bangkok, Thailand, Cape Town, South Africa, and New York City, USA.

Twenty-six previously unreported genetic missense variants were detected in 24 individuals out of 505 individuals sequenced for the coding DNA reference sequence NM_001824.4. All individuals that had detectable missense variants were heterozygous for those variants. One participant from Bangkok had two missense variants and one participant from Cape Town had two missense variants. Across the three geographic locations evaluated for genetic variants for this kinase, no *CKM* variants were found in more than one geographic region.

Investigation of sequence variants of *PKM* detected across three geographic locations. In PBMC and vaginal tissue, *PKM* has been shown to phosphorylate TFV-MP to TFV-DP. We employed a targeted next-generation sequencing assay to investigate the existence of genetic variants in *PKM* found in the exon regions of this kinase. For the DNA reference sequence corresponding to *PKM* (NM_001206796.1), 39 total variants comprised of SNVs and deletions that are predicted to be reflected in the amino acid sequence were detected across 33 individuals. The 39 observed variants, which include both previously reported and previously unreported variants, are listed in Supplementary Table 2. All 33 individuals carrying variants in *PKM* were observed to be heterozygous for all the described *PKM* variants. Of the 33 individuals observed to have detectable SNVs or deletions, 11 individuals were located at the Bangkok study site, 10 individuals were located at the Cape Town study site, and 12 individuals were located at the New York City study site. One individual at the Cape Town study site was observed to carry two variants. Two individuals at the New York City study site were observed to carry three variants. Thirty-three previously unreported single nucleotide variations or base deletions were detected across 29 individuals (Table 4). Three reported variants, found in the dbSNP, were detected and are listed as follows: rs141732747 (NM_001206796.1:c.827C>A) was detected in one individual from New York City; rs185164430 (NM_001206796.1:c.1695C>G) was detected in one individual from New York City and rs778625515 (NM_001206797.1:c.132C>G) was detected in one individual from Cape Town. For the variants detected in *PKM*, functional predictions at the protein level were unavailable because the *in silico* tools, SIFT and Polyphen, were

unable to achieve sufficient sequence diversity information in the multiple alignments necessary for prediction.

Geographic Location	Variant (ref.>alt.)	cDNA Position	Coding DNA Sequence Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon
Bangkok	A>G	766	367	123	C>R	3/12
Bangkok	T>C	904	505	169	T>A	5/12
Bangkok	G>A	1007	608	203	T>I	6/12
Bangkok	T>C	1058	659	220	N>S	6/12
Bangkok	C>G	1399	1000	334	E>Q	7/12
Bangkok	C>T	1502	1103	368	R>H	8/12
Bangkok	G>A	1667	1268	423	A>V	9/12
Bangkok	C>T	1752	1353	451	M>I	9/12
Bangkok	T>C	1837	1438	480	T>A	10/12
Bangkok	A>G	2065	1666	556	W>R	11/12
Cape Town	G>A	437	38	13	T>M	1/12
Cape Town	G>C	617	218	73	A>G	3/12
Cape Town	C>T	688	289	97	A>T	3/12
Cape Town	A>T	850	451	151	S>T	4/12
Cape Town	T>C	1027	628	210	K>E	6/12
Cape Town	T>C	1412	1013	338	N>S	7/12
Cape Town	T>C	1762	1363	455	I>V	10/12
Cape Town	G>A	1792	1393	465	H>Y	10/12
Cape Town	C>T	2122	1723	575	G>S	12/12
Cape Town	G>T	2177	1778	593	S>Y	12/12
New York City	G>A	430	31	11	L>F	1/12
New York City	C>T	1345	946	316	A>T	7/12
New York City	T>C	1475	1076	359	E>G	8/12
New York City	T>A	1492	1093	365	M>L	8/12
New York City	T>C	1618	1219	407	S>G	9/12
New York City	C>T	1655	1256	419	G>D	9/12
New York City	T>A	1712	1313	438	E>V	9/12
New York City	C>A	1741	1342	448	A>S	9/12
New York City	A>T	1844	1445	482	L>H	10/12
New York City	C>T	1849	1450	484	E>K	10/12
New York City	A>G	1979	1580	527	V>A	11/12
New York City	C>T	2047	1648	550	D>N	11/12

Table 4. Previously unreported *PKM* missense variants detected in clinical trial participants located in Bangkok, Thailand, Cape Town, South Africa, and New York City, USA.

Thirty-three previously unreported genetic missense variants were detected in 29 individuals out of 505 individuals sequenced for the coding DNA reference sequence NM_001206796.1. All individuals that had detectable variants were heterozygous for those variants. Across the three geographic locations evaluated for genetic variants for this kinase, no *PKM* variants were found in more than one geographic region. Using *in silico* tools PolyPhen and SIFT, no functional predictions were available for *PKM*.

Investigation of sequence variants of *PKLR* detected across three geographic locations. *PKLR* has been shown to phosphorylate TFV-MP to TFV-DP in PBMC and vaginal tissue, leading us to investigate the existence of genetic variants in *PKLR* predicted to yield functional effects on this kinase. For the DNA reference sequence corresponding to *PKLR* (NP_000289.1), 39 total variants comprised of SNVs and deletions that are predicted to be reflected in the amino acid sequence were detected across 35 individuals. These observed variants are listed in Supplementary Table 2. Thirty-four individuals were determined to be heterozygous for the variants detected, with one individual identified as homozygous for one variant. Of the 34 individuals observed to have detectable SNVs or deletions, 10 individuals were located at the Bangkok study site, 11 individuals were located at the Cape Town study site, and 14 individuals were located at the New York City study site. Two individuals at the Cape Town study site were observed to be carrying two variants compared to 4 individuals at the New York City study site that carried two variants. Thirty-one previously unreported single nucleotide variations or base deletions were detected across 26 individuals (Table 5). One individual from Cape Town was homozygous for the variant G180V. All other participants that showed detected missense variants were heterozygous for the variants. One individual from Cape Town carried two missense variants and four individuals from

New York City had two missense variants. Across the three geographic locations evaluated for genetic variants in this kinase, no PKLR variants were found in more than one geographic region, however, in addition only two of the same variants were detected in more than one individual. Of the 505 individuals sequenced for PKLR missense variants, nine individuals were observed to have at least one predicted "deleterious" and "probably damaging" variant, yielding a frequency of 1.8% (9/585). *In silico* tools SIFT and PolyPhen were used to predict the functional consequences of these variants. Six reported variants, found in the dbSNP, were detected and are listed as follows: rs147689373 (NM_000298.5:c.829G>A) was found in two individuals from New York City; rs116100695 (NM_000298.5:c.1456C>T) was found in one individual from New York City and one individual from Cape Town; rs189360283 (NM_000298.5:c.1345C>T) was found in one individual from Cape Town; rs201979697 (NM_000298.5:c.907C>T) was found in one individual from Cape Town; rs201406712 (NM_000298.5:c.1435C>T) was found in one individual from New York City and rs150077703 (NM_000298.5:c.92C>A) was found in one individual from Bangkok.

Geographic Location	Variant (ref.>alt.)	cDNA Position	Coding DNA Sequence Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon	SIFT Prediction	PolyPhen Prediction
Bangkok	G>A	89	50	17	S>F	1/11	deleterious(0)	benign(0)
Bangkok	A>T	106	67	23	L>I	1/11	tolerated(0.06)	benign(0.024)
Bangkok	C>T	161	122	41	R>Q	2/11	deleterious(0.01)	benign(0.175)
Bangkok	G>C	302	263	88	T>S	2/11	tolerated(0.16)	possibly_damaging(0.793)
Bangkok	G>C	368	329	110	A>G	3/11	deleterious(0)	possibly_damaging(0.656)
Bangkok	A>T	1240	1201	401	C>S	8/11	deleterious(0.02)	possibly_damaging(0.837)
Bangkok	A>G	1334	1295	432	V>A	9/11	deleterious(0)	benign(0.062)
Bangkok	T>C	1652	1613	538	E>G	10/11	tolerated(0.11)	benign(0.024)
Cape Town	C>T	375	336	112	M>I	3/11	deleterious(0.01)	probably_damaging(1)
Cape Town	C>A	578	539	180	G>V	5/11	deleterious(0)	probably_damaging(0.996)
Cape Town	G>T	947	908	303	P>Q	6/11	deleterious(0.05)	benign(0.13)
Cape Town	C>T	1033	994	332	G>S	7/11	deleterious(0.01)	probably_damaging(0.995)
Cape Town	A>G	1040	1001	334	M>T	7/11	deleterious(0.01)	probably_damaging(0.995)
Cape Town	A>T	1091	1052	351	L>Q	7/11	deleterious(0)	probably_damaging(0.992)
Cape Town	G>A	1313	1274	425	A>V	9/11	deleterious(0.01)	benign(0.111)
Cape Town	C>T	1343	1304	435	R>Q	9/11	tolerated(0.63)	benign(0.01)
Cape Town	T>A	1465	1426	476	T>S	9/11	tolerated(0.23)	benign(0.002)
Cape Town	A>G	1576	1537	513	F>L	10/11	tolerated(0.25)	possibly_damaging(0.624)
New York City	G>C	197	158	53	T>S	2/11	tolerated(0.47)	benign(0.008)
New York City	G>T	368	329	110	A>D	3/11	deleterious(0)	possibly_damaging(0.9)
New York City	A>G	460	421	141	F>L	4/11	tolerated(0.45)	benign(0.082)
New York City	G>A	473	434	145	P>L	4/11	tolerated(0.1)	possibly_damaging(0.653)
New York City	A>T	524	485	162	I>N	4/11	deleterious(0)	probably_damaging(0.999)
New York City	G>A	530	491	164	T>I	4/11	deleterious(0.01)	probably_damaging(1)
New York City	A>G	580	541	181	S>P	5/11	tolerated(0.26)	benign(0.315)
New York City	A>T	866	827	276	V>E	6/11	deleterious(0)	probably_damaging(0.986)
New York City	A>C	1230	1191	397	D>E	8/11	deleterious(0.01)	possibly_damaging(0.872)
New York City	C>T	1234	1195	399	A>T	8/11	tolerated(0.29)	probably_damaging(0.94)
New York City	G>A	1448	1409	470	A>V	9/11	deleterious(0.02)	probably_damaging(0.995)
New York City	A>G	1520	1481	494	I>T	10/11	deleterious(0)	probably_damaging(0.985)
New York City	A>G	1583	1544	515	L>S	10/11	deleterious(0.03)	probably_damaging(0.97)

Table 5. Previously unreported *PKLR* missense variants detected in clinical trial participants located in Bangkok, Thailand, Cape Town, South Africa, and New York City, USA.

Thirty-one previously unreported genetic missense variants were detected in 26 individuals out of 505 individuals sequenced for the coding DNA reference sequence NP_000289.1.

Discussion

This study identified 103 previously unreported genetic variants of the kinases *AK2, CKM, PKM,* and *PKLR* in 505 clinical trial participants located in Thailand, Cape Town, and the United States. To increase confidence in the variants identified, our analyses conformed to the American College of Medical Genetics and Genomics clinical laboratory standards for next-generation sequencing [8]. This method of analysis has been previously described [7]. To compare frequencies in this study of previously reported genetic variants, data from the 1000 Genomes Project [9] and the Exome Aggregation Consortium (ExAC), which includes over 60,000 genotyped, unrelated individuals, [12] was used because of the population and, in the case of the 1000 Genomes Project, location information these datasets provide.

In comparison with a previous analysis from our laboratory that sequenced 142 women from the geographic regions Durban, South Africa, Kampala, Uganda, and four locations within the United States [7], we detected one overlapping variant for the four kinases sequenced. Interestingly, this genetic variant for *PKM* SNP rs778625515 (NM_001206796.1:c.354C>G) was detected in one clinical trial participant at the USA study site from our previous analysis and shared with one participant at the Cape Town study site in our current study. At the protein level, this *PKM* variant is predicted to result in a mutation from asparagine to lysine at amino acid residue 118 (N118K). This mutation is located adjacent to a beta strand comprised of amino acids 119 to 121 [10]. Because of its location two amino acids away from a reported ATP binding site, it is possible binding of this substrate will be disrupted, affecting kinase function; however, as noted above, the gene encoding *PKM* currently has no functional prediction of this

variant using *in silico* tools PolyPhen or SIFT. Functional predictions of the variants detected in *PKM* at the protein level were unavailable because the *in silico* tools used for prediction, SIFT and PolyPhen, were not supplied with sufficient sequence diversity required for the performed multiple sequence alignments necessary for the algorithms used in these predictions.

Continuing the comparison to our previous study [7], the SNV rs147689373 (NM_000298.5:c.829G>A) in the *PKLR* kinase was also detected in two individuals in the New York City study site we investigated. This yielded a 1.3% (2/149) frequency within New York City and a 0.40% (2/505) frequency in our study overall. This variant was also detected in one participant in South Africa in our previous study [7]. In the 1000 Genomes Project, the overall frequency for this variant was 0.26%. The only location in the USA included in the 1000 Genomes Project reporting frequency was 0.78% in the Mexican Ancestry from Los Angeles USA. Remaining locations contributing to the overall frequency listed for this variant in the 1000 Genomes Project were 2.1% in African Caribbeans in Barbados, 1.0% Esan in Nigeria, 1.52% Luhya in Webuya Kenya, 1.18% Mende in Sierra Leone, and 0.46% Yoruba in Ibadan, Nigeria. At the protein level, this variant results in a mutation at the 277th amino acid with a substitution from glutamic acid to lysine. Other previously reported variants around this region of the protein have been associated with pyruvate kinase deficiency [11].

Of the previously reported variants detected in our study, the SNP in *CKM* rs11559024 (NM_001824.4:c.248A>G) was the only variant reported at an overall frequency greater than 1% across the ExAC database. This variant was detected in one individual from New York City, exhibiting a 0.20% (1/505) overall frequency across the three geographic locations in this study and a 0.67% (1/149) frequency in the New York City population we examined. Across data listed in the ExAC database, a total frequency of 1.1% throughout all the subpopulations included in the database is reported. All subpopulations exhibited frequencies of this variant, with Finnish European at 1.94%, non-Finnish European at 1.51%, South Asian at 0.58%, Other at 0.55%, Latino at 0.36%, African at 0.36%, and the East Asian population exhibiting 0.01% frequency. Consulting the 1000 Genomes Project, this variant is reported at the following frequencies in the following sites located in the USA: 0.25% in Utah Residents with Northern and Western European Ancestry, 0.49% in Gujarati Indians from Houston, Texas population, and 0.78% in individuals reporting Mexican Ancestry from Los Angeles USA. Populations from the 1000 Genomes Project outside the United States showed frequencies at 0.58% in Bengali from Bangladesh, 0.49% in Han Chinese in Beijing, 0.53% in Colombians from Medellin, Colombia, 1.52% in Finnish in Finland, 1.65% in British in England and Scotland, 1.04% in Punjabi from Lahore, Pakistan, and 0.48% in Puerto Ricans from Puerto Rico. Overall, in the populations included in the 1000 Genomes Project combined, the frequency of this variant was 0.38%-- a higher overall frequency than that observed in our study. The resulting protein from this variant would exhibit a residue change at the 83rd amino acid of the protein, resulting in the mutation of a glutamic acid to glycine.

Using *in silico* functional prediction tools, this substitution is predicted to be "deleterious" using the SIFT algorithm and "possibly damaging" using PolyPhen.

Possible effects on TFV activation by these variants can be envisioned by examining the resulting protein sequences. One example of a previously unreported deleterious and possibly damaging *AK2* variant was found at the Cape Town study site, predicted to result in an amino acid substitution resulting in the mutation K28R. The protein position at 28 is reported to be in the middle of the ATP binding region of *AK2* [13]. This kinase, reported to catalyze the transfer of a phosphate group from an ATP molecule, would be functionally affected by this variant by having a disrupted source of phosphate to transfer onto TFV. This potential for binding disruption could impact TFV phosphorylation.

The fact that we observed very little overlap between individuals in the variants identified suggests that while genetic differences in these kinases exist, the variants are of low frequency making the probability of detecting mutations that might impact TFV activation in a given population difficult to predict. Thus, genetic differences in TFV activation may be independent of one's ethnicity or geographic origin. Nevertheless, our results offer new, previously unreported variants across the *AK2, CKM, PKM,* and *PKLR* genes, which encode the kinases that play a role in tenofovir activation. Of note, several of these kinases were predicted to result in a loss or a decrease in the function of the protein. Taken together, this work lends further evidence to the possibility that genetic

variation in TFV activation may underlie clinically observed differences in efficacy that cannot be explained by adherence alone.

References

[1] Schooley, R.T., Ruane, P., Myers, R.A., Beall, G., Lampiris, H., Berger, D., et al., 2002. Tenofovir DF in antiretroviral-experienced patients: results from a 48-week, randomized, double-blind study. AIDS 16 (9), 1257–1263.

[2] USFDA, FDA Approves First Medication to Reduce HIV Risk., FDA Consumer Health Information, 2012.

[3] Grant, R.M., Lama, J.R., Anderson, P.L., McMahan, V., Liu, A.Y., Vargas, L., et al., 2010. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N. Engl. J. Med. 363 (27), 2587–2599.

[4] Thigpen, M.C., Kebaabetswe, P.M., Paxton, L.A., Smith, D.K., Rose, C.E., Segolodi, T.M., et al. 2012. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. N. Engl. J. Med. 367 (5) 423-434.

[5] Baeten, J.M., Donnell, D., Ndase, P., Mugo, N.R., Campbell, J.D., Wangisi, J., et al., 2012. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N. Engl. J. Med. 367 (5), 399–410.

[6] Robbins, B.L., Greenhaw, J., Connelly, M.C., Fridland, A., 1995. Metabolic pathways for activation of the antiviral agent 9-(2-phosphonylmethoxyethyl)adenine in human lymphoid cells. Antimicrob. Agents Chemother. 39 (10), 2304–2308.

[7] Lade, J.M., To, E.E., Hendrix, C.W., Bumpus, N.N., 2015. Discovery of Genetic Variants of the Kinases That Activate Tenofovir in a Compartment-specific Manner. EBioMedicine. 2 (9), 1145–1152.

[8] Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., et al., 2013. ACMG clinical laboratory standards for next-generation sequencing. Genet. Med. 15 (9), 733–747.

[9] Abecasis, G.R., Auton, A., et al. 2015. A global reference for human genetic variation. Nature. 526 (68 – 74).

[10] Anastasiou, D., Yu, Y., Israelsen, W.J., Jiang, J.K., Boxer, M.B., Hong, B.S., Tempel, W., et al. 2012. Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nature Chemical Biology. 8 (10) 839-847.

[11] van Wijk, R., Huizinga, E.G., van Wesel, A.C., van Oirschot, B.A., Hadders, M.A., van Solinge, W.W., 2009. Fifteen novel mutations in PKLR associated with

pyruvate kinase (PK) deficiency: structural implications of amino acid substitutions in PK. Hum. Mutat. 30 (3) 446-453.

[12] Exome Aggregation Consortium, et al. 2015. Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv. Cold Spring Harbor Laboratories.

[13] Bunkoczi, G., Filippakopoulos, P., Debreczeni, J.E., Turnbull, A., Papagrigoriou, E., Savitsky, P., et al. 2006. Structure of Human Adenylate Kinase 2. PD ID: 2C9Y, 2006.

Chapter 3:

Identification of genetic variants discovered in candidate kinases involved in phosphorylation of tenofovir and emtricitabine.

Abstract

Tenofovir (TFV) and emtricitabine (FTC) are the two components that make up the only drug currently approved by the FDA for HIV pre-exposure prophylaxis (PrEP). TFV and FTC are nucleoside mimics that must be phosphorylated to the nucleotide triphosphate analogs that competitively inhibit HIV reverse transcriptase. TFV activation by kinases adenylate kinase 2 (AK2), creatine kinase, muscle (CKM), pyruvate kinase, muscle (PKM) and pyruvate kinase, liver and red blood cell (PKLR) has been demonstrated in peripheral blood mononuclear cells (PBMC), vaginal, and colorectal tissue, indicating compartment-specific phosphorylation of this compound. FTC requires three phosphorylation steps to form emtricitabine-triphosphate (FTC-TP), with deoxycytidine kinase (DCK) and thymidine kinase 1 (TK1) proposed to phosphorylate FTC to emtricitabine-monophosphate (FTC-MP). Cytidine monophosphate kinase 1 (CMPK1) is suggested to phosphorylate FTC-MP to emtricitabine diphosphate (FTC-DP). Finally, phosphoglycerate kinase 1 (PGK1) is identified as a candidate for FTC-DP phosphorylation to the nucleotide triphosphate analog FTC-TP, the pharmacologically active competitive inhibitor of HIV reverse transcriptase. In this study, we apply next-generation sequencing toward identifying genetic variants in the genes encoding AK2, CKM, PKM, PKLR, DCK, TK1, CMPK1, and PGK1 in clinical samples from the HIV Prevention Trials Network study 069. We identified 17 previously unreported genetic variants of these kinases. Using bioinformatics tools, se also put forth predictions of the functional impact of these genetic variants. Building upon previous studies, our results further push forward the concept that genetic variation may contribute to inter-individual variability in TFV and FTC activation.

Introduction

The FDA's approval of the combination antiretroviral therapy regimen Truvada ® for use as HIV pre-exposure prophylaxis (PrEP) in July 2012 remains a capstone in HIV prevention strategy as the first approved pharmacological approach for preventing HIV infection in HIV uninfected individuals [1]. Playing an integral role toward FDA approval of Truvada ® as PrEP, the clinical trial iPrEX demonstrated 92% of men who have sex with men (MSM) or transgender women (TGW) were protected from HIV infection when adherent to the drug regimen [2]. PrEP as a strategy for HIV prevention was further supported by the Botswana TDF2 study, demonstrating 78% of HIV-uninfected men and women were protected against HIV infection [3], along with the Partners PrEP study, demonstrating 90% protection in sero-discordant couples [4]. In all of these studies, even when adherence was taken into account, full protection against HIV was not observed. We have previously put forth data through which we propose that this could occur at least in part, due to genetic variations in the activation of tenofovir (TFV), a component of the PrEP regimen Truvada ®, by kinases expressed in peripheral blood mononuclear cells (PBMC), vaginal tissue, and colorectal tissue [5].

Truvada ® is comprised of TFV, prescribed as the prodrug tenofovir disoproxil fumarate, and emtricitabine (FTC). Both compounds are nucleoside or nucleotide reverse transcriptase inhibitors (NRTIs). TFV is an adenosine monophosphate analog and FTC is a cytidine analog [6]. Both compounds consequently require intracellular phosphorylation to form their active nucleotide triphosphate analogs. To form the nucleotide triphosphate analog that competitively inhibits HIV reverse transcriptase, TFV must be phosphorylated twice. We have demonstrated adenylate kinase 2 (AK2) phosphorylation of TFV to tenofovir-monophosphate (TFV-MP) in PBMC, vaginal tissue, and colorectal tissue [5]. Interestingly, our lab found that phosphorylation of TFV-MP is compartment specific in that creatine kinase, muscle (CKM) can phosphorylate TFV-MP to TFV-DP in colon tissue, while pyruvate kinase, muscle (PKM) and pyruvate kinase,

liver and red blood cell (PKLR) were observed to catalyze the formation of TFV-DP in PBMC and vaginal tissue [5]. Like TFV, FTC is also a NRTI that must be phosphorylated intracellularly to form FTC-TP. However, unlike the monophosphate analog TFV which needs to be phosphorylated twice, FTC requires three phosphorylation steps [7]. Consulting previous studies, we suggest four candidate kinases that could activate FTC in HIV susceptible tissues. Deoxycytidine kinase (DCK) has been shown *in vitro* to phosphorylate FTC to FTC-MP using calf thymus deoxycytidine kinase [7]. Another candidate kinase for phosphorylating FTC to FTC-MP, thymidine kinase 1 (TK1), has been shown to phosphorylate zidovudine and stavudine [8]. Although zidovudine and stavudine are thymidine analogs rather than a cytidine analog, both are dideoxynucleosides as FTC is, with all compounds lacking both 2'- and 3'-hydroxyl groups in their sugar ring. Phosphorylation of FTC-MP to FTC-DP is hypothesized due to demonstrated activity of the kinase cytidine monophosphate kinase 1 (CMPK1) toward lamivudine monophosphate (3TC-MP) [9]. Lamivudine (3TC) is a cytidine analog structurally similar to emtricitabine, with the only difference between the two being a fluorine atom in emtricitabine at the 5 position of the cytidine base. A purified protein assay exhibited phosphorylation of 3TC-MP to lamivudine diphosphate (3TC-DP) using purified human CMPK1. Finally, to form the pharmacologically active metabolite FTC-TP, we predict phosphorylation of FTC-DP to FTC-TP by phosphoglycerate kinase 1 (PGK1). The basis for this prediction stems from previously demonstrated phosphorylation of the diphosphorylated anabolite of the deoxynucleoside analog L-Fd4C [10]. This compound is structurally similar to FTC, except for a double bond between the pentose 2' and 3' positions rather than a sulfur atom at the 3' position that FTC exhibits. PGK1 was purified from HepG2 cells and used to investigate the presence of its phosphorylation activity toward L-Fd4C-DP.

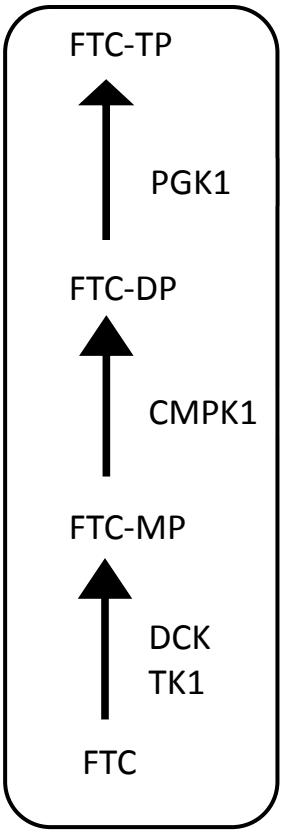

Figure 1. Schematic showing proposed activation of FTC by nucleotide kinases. DCK and TK1 are hypothesized to phosphorylate FTC to FTC-MP. CMPK1 is suggested to phosphorylate FTC-MP to FTC-DP. To form the nucleotide triphosphate analog FTC-TP, PGK1 is postulated to phosphorylate FTC-DP.

Along with identifying TFV-activating kinases in PBMC, vaginal tissue, and colorectal tissue, previous work in our lab also put forth identification of previously unreported genetic variants in the kinases AK2, CKM, PKM, and PKLR [5]. Variants discovered in the previous study were predicted to have an impact on the function of the protein sequenced, exhibiting deleterious and damaging phenotype predictions. These functional predictions that could negatively affect TFV activation may provide an explanation for the incomplete protection against HIV despite adherence to PrEP.

Our present study builds further upon the idea of genetic variation in drug activation through nucleotide kinase variants by sequencing clinical trial participants not only for the TFV-

phosphorylating kinases but also for the genes that encode DCK, TK1, CMPK1, and PGK1. Investigating the existence of genetic variants in these kinases could contribute to differences in HIV protection due to differences in inter-individual FTC phosphorylation. To test for the presence of such variants, we applied a targeted next-generation sequencing assay to sequence the genes *AK2, CKM, PKM, PKLR, DCK, TK1, CMPK1,* and *PGK1* using genomic DNA isolated from whole blood. The clinical trial participants in this study were from the HIV Prevention Trials Network study HTPN 069. This work identifies previously unreported genetic variants in kinases that have been shown to activate TFV as well as in kinases predicted to activate FTC. Using *in silico* functional prediction tools, these variants were investigated for any possible functional effects on the expressed protein. Given the presence of variants that were predicted to impede protein function, it is possible that TFV phosphorylation and therefore activation becomes compromised in individuals carrying these variants. These findings expand the current knowledge of genetic variants that may contribute to differences in HIV protection, putting forward nucleotide kinases that may be involved in FTC activation and contributing elements other than adherence that may govern PrEP efficacy.

Materials and Methods

Clinical study sites and sample collection. Whole blood was obtained from HIV-uninfected individuals (n = 498) enrolled in the HIV Prevention Trials Network study HPTN 069 across 13 clinical research sites (CRS) located within the United States: Bridge HIV CRS, San Francisco, CA (UCSF); Case CRS, Cleveland, OH (Case); Chapel Hill CRS, Chapel Hill, NC (UNC); Cornell CRS, New York, NY (Cornell); Fenway Health CRS, Boston, MA (Fenway); George Washington University CRS, Washington, DC (GW); Hospital of University of Pennsylvania, Philadelphia, PA (UPenn); Johns Hopkins University CRS, Baltimore, MD (JHU); New Jersey Medical School CRS, Newark, NJ (UMDNJ); Pitt CRS, Pittsburgh, PA (UPitt); Puerto Rico AIDS CRS, San Juan, PR (UPR); UCLA Care Center CRS, Los Angeles, CA (UCLA);

University of Washington AIDS CRS, Seattle, WA (UW). All participants sequenced consented to genetic testing. Information about participants genotyped in this study is summarized in Table 1. These participants were MSM or TGW and women who have sex with men (WSM) in locations across the United States.

HPTN 069 Participants, n = 498		
Study Site	Participant Sex	n
Bridge HIV CRS, San Francisco, CA	Female	0
	Male	48
Case CRS, Cleveland, OH	Female	16
	Male	33
Chapel Hill CRS, Chapel Hill, NC	Female	12
	Male	28
Cornell CRS, New York, NY	Female	11
	Male	22
Fenway Health CRS, Boston, MA	Female	13
	Male	43
George Washington University CRS, Washington, DC	Female	19
	Male	44
Hospital of University of Pennsylvania, Philadelphia, PA	Female	25
	Male	11
Johns Hopkins University CRS, Baltimore, MD	Female	0
	Male	21
New Jersey Medical School CRS, Newark, NJ	Female	26
	Male	0
Pitt CRS, Pittsburgh, PA	Female	5
	Male	23
Puerto Rico AIDS CRS, San Juan, PR	Female	14
	Male	21
UCLA Care Center CRS, Los Angeles, CA	Female	2
	Male	28
University of Washington AIDS CRS, Seattle, WA	Female	7
	Male	26

Table 1. Location and sex information of HPTN 069 participants who consented to genotyping.

Genomic DNA isolation from HPTN 069 whole blood samples. Genomic DNA was isolated from 200 µL of whole blood from each individual using a QIAamp 96 DNA Blood Kit (QIAGEN, Valencia, CA). Genomic DNA was extracted following protocol "Purification of DNA from Whole Blood,

Plasma, Serum, or Body Fluids – (EN)" from QIAGEN. Purified DNA was eluted using 150 µL DEPC-treated, nuclease-free water (Quality Biological, Inc., Gaithersburg, MD) and concentrated using a ZR-96 DNA Clean-up KitTM (Zymo Research, Irvine, CA). Resulting concentrated and

purified genomic DNA was eluted in 12 μL DEPC-treated, nuclease-free water.

Next-generation sequencing target design, sample preparation, and analysis. Sequencing of the kinases that have been show to phosphorylate TFV to TFV-DP AK2, CKM, PKM and PKLR was executed using the Illumina TruSeq Custom Amplicon kit v1.5 (Illumina, San Diego, CA). To sequence the kinases proposed to phosphorylate FTC to FTC-TP, CMPK1, DCK, PGK1 and TK1, the Illumina TruSeq Custom Amplicon kit v1.5 was also used. Probes for sequencing the combination of these seven kinases were designed using Illumina DesignStudio software as previously described [5]. Genomic DNA isolated from clinical samples was processed following the Illumina TruSeq Custom Amplicon Library Preparation Guide (Part Number 15027983 Rev. C, August 2013). DNA concentration was measured using a Qubit ® 3.0 Fluorimeter (Thermo Scientific, New York, NY). A plasmid containing a known sequence for AK2 was used as an additional sequencing control. Fifty ng of DNA input were used per DNA sample sequenced. The resulting prepared DNA library (6 μL) was diluted in 594 μL HT1 buffer containing 1% PhiX sequencing control. Illumina VariantStudio software was used to annotate and analyze variant read quality as previously described [5].

Results

Analysis of individuals carrying genetic variants in tenofovir-activating kinases *AK2, CKM, PKM,* and *PKLR*. A targeted assay to sequence the exonic regions of the genes *AK2, CKM, PKM*, and *PKLR* was applied to investigate the presence of genetic variants in these nucleotide kinases. Across the United States, 13 clinical research sites recruited participants, resulting in 498 total individuals (n = 498) who consented to genetic testing. Of these individuals, nine were observed to carry a single nucleotide variant in the *AK2* gene, yielding a 1.8% frequency (9/498 individuals) of *AK2* variants within this population. One individual carrying one variant that was

predicted to be deleterious (0.2%, 1/498 individuals) was detected among the nine individuals with observed *AK2* variants. In the *CKM* gene, 8 individuals were found to carry variants, showing 1.6% (8/498 individuals) of individuals carrying variants in this kinase. Of these individuals, five were predicted to carry deleterious variants, yielding a frequency of 1% (5/498 individuals). The *PKM* gene sequenced across all individuals detected 11 participants to have variants (2.2% frequency, 11/498 individuals). The lack of sequence diversity required to perform the sequence alignments for this kinase did not allow for functional predictions using the *in silico* tools SIFT and Polyphen. Sequencing the kinase *PKLR* yielded 11 individuals predicted to carry genetic variants resulting in amino acid sequence alterations. All 11 individuals were observed to carry variants that were predicted to have a deleterious functional impact. These 11 individuals make up the observed frequency of 2.2% (11/498 individuals). Variant distributions of the tenofovir-activating kinases *AK2, CKM, PKM*, and *PKLR* in individuals are demonstrated using a Venn diagram in Figure 2.

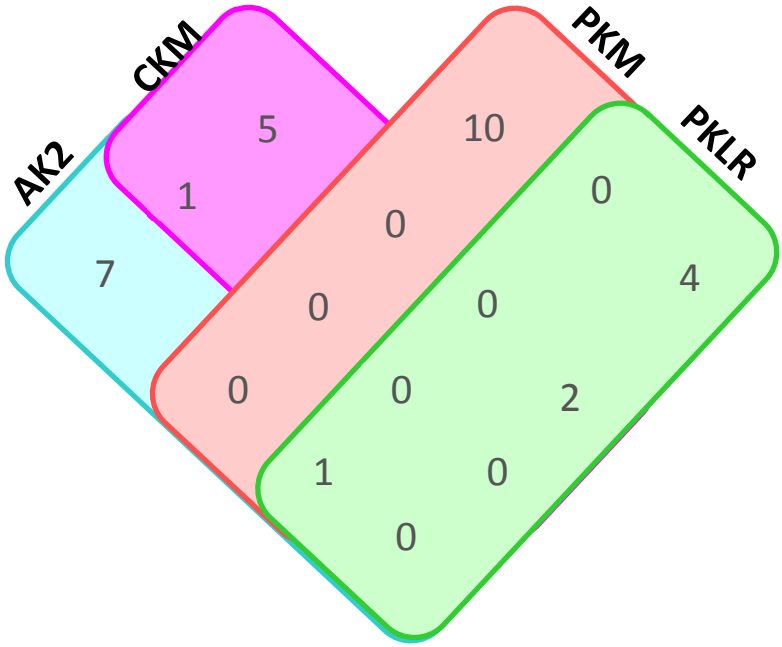

Figure 2. Distribution of genetic variants in TFV-activating kinases across 30 clinical trial participants. Each rectangle represents a nucleotide kinase that was sequenced, with *AK2* in blue, *CKM* in pink, *PKM* in orange, and *PKLR* depicted in green. Numerical values indicate the number of individuals detected to carry a single nucleotide variation

or deletion. Overlapping regions of each rectangle indicate the number of individuals with genetic variants in more than one kinase.

Analysis of *AK2* genetic variants detected in nine clinical trial participants. In previous work, our laboratory has demonstrated *AK2* to phosphorylate TFV to TFV-MP in PBMC, colon tissue, and vaginal tissue [5]. For the DNA reference sequence corresponding to *AK2* (NM_001625.3), we observe two previously unreported variants and three previously reported variants. These five variants were detected across nine individuals (1.8% frequency, 9/498 individuals). Variants observed are shown in Table 2 and detailed in Supplementary Table 1. All individuals with detected variants were observed to be heterozygous for these variants. Both unreported variants were predicted to be tolerated, with one variant resulting in a N157S mutation and the other resulting in a P131T mutation at the protein level. Three variants previously listed in the Single Nucleotide Polymorphism database (dbSNP) were detected and are listed as follows: rs138577419 (NM_001625.3:c.49C>G) was detected in one individual, rs12116440 (NM_001625.3:c.625G>A) was detected in two individuals, with the variant rs61750965 (NM_001625.3:c.386G>A) also observed in two individuals. Of the individuals exhibiting genetic variants for *AK2*, only one individual was observed to carry a predicted deleterious variant, resulting in a 0.2% frequency.

Variant (ref.>alt.)	cDNA Position	Coding DNA Sequence Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
T>C	553	470	157	N>S	5/6	tolerated(0.15)	benign(0.001)		0.2% (1/498)
G>C	132	49	17	R>G	1/6	deleterious(0)	possibly_damaging(0.514)	rs138577419	0.2% (1/498)
C>T	708	625	209	A>T	6/6	tolerated(0.5)	benign(0.056)	rs12116440	0.4% (2/498)
C>T	469	386	129	S>N	4/6	tolerated(0.49)	benign(0.001)	rs61750965	0.4% (2/498)
G>T	474	391	131	P/T	4/6	tolerated(0.07)	benign(0.037)		0.6% (3/498)

Table 2. *AK2* missense variants detected in nine clinical trial participants.

Identification of genetic variants in *CKM* gene yields variants in eight clinical trial participants. TFV-MP must be phosphorylated to TFV-DP to become the active competitive inhibitor of HIV reverse transcriptase. We have demonstrated that in colon tissue, *CKM* phosphorylates TFV-MP to TFV-DP, generating the pharmacologically active reverse transcriptase inhibitor. Comparison to the DNA reference sequence for *CKM* (NM_001824.4), through sequencing the exon region of this gene yielded four previously unreported variants and two variants found in the SNP database. These six variants were detected across eight individuals, presenting a 1.6% (8/498) frequency of individuals carrying variants. Further details regarding these variants are in Table 3 and Supplementary Table 2. All individuals with *CKM* genetic variants were observed to be heterozygous for those variants. The four previously unreported variants resulted in a predicted deleterious T35I protein mutation, a predicted deleterious T53N mutation, and two predicted tolerated mutations at R43Q and G4C. Of the six variants detected for *CKM* in this population, two were found in dbSNP. The variant rs149354459 (NM_001824.4:c.752G>C) was observed in one individual, while the variant rs11559024 (NM_001824.4:c.248A>G) was observed in two individuals. Across all variants detected in the *CKM* gene sequenced, five individuals exhibited variants that were predicted to be deleterious (1% frequency, 5/498).

Variant (ref.>alt.)	cDNA Position	CDS Position	Protein Position	Amino Acid Substitution (ref.>alt.)	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
G>A	279	104	35	T/I	2/8	deleterious(0)	probably_damaging(0.914)		0.2% (1/498)
C>G	927	752	251	R/P	6/8	tolerated(0.09)	benign(0.129)	rs149354459	0.2% (1/498)
T>C	423	248	83	E/G	3/8	deleterious(0.01)	possibly_damaging(0.467)	rs11559024	0.6% (3/498)
C>T	303	128	43	R/Q	2/8	tolerated(0.07)	benign(0.288)		0.2% (1/498)
C>A	185	10	4	G/C	2/8	tolerated(0.1)	probably_damaging(0.998)		0.2% (1/498)
G>T	330	155	52	T/N	2/8	deleterious(0.04)	probably_damaging(0.972)		0.2% (1/498)

Table 3. *CKM* missense variants detected in eight clinical trial participants.

***PKM* genetic variants observed in 11 clinical trial participants.** Our laboratory has previously demonstrated phosphorylation of TFV-MP to TFV-DP by *PKM* in PBMC and vaginal tissue [5]. In the 498 participants sequenced for the DNA reference sequence corresponding to *PKM* (NM_001206796), 11 individuals were found to carry variants at the amino acid level (2.2% frequency, 11/498). Across these 11 individuals, five total variants were observed, with three previously unreported variants and two variants found in the SNP database. These variants are presented in Table 4 and detailed in Supplementary Table 3. Variants detected in *PKM* did not have functional predictions at the protein level because the *in silico* prediction tools PolyPhen and SIFT did not have sufficient sequence diversity information in the multiple alignments used to predict variant functional impact. The two previously reported genetic variants for *PKM* were rs180716407 (NM_001206796.2:c.14C>G) and rs201533100 (NM_001206796.2:c.395T>C). The variant rs180716407 was detected in five individuals while the variant rs201533100 was detected in one individual.

Variant (ref.>alt.)	cDNA Position	CDS Position	Protein Position	Amino Acids	Exon	dbSNP ID	Variant Frequency
G>G/T	931	532	178	L/I	5/12		0.4% (2/498)
G>G/C	413	14	5	S/*	1/12	rs180716407	0.6% (3/498)
T>T/A	505	106	36	T/S	2/12	rs147562047	0.2% (1/498)
A>A/G	794	395	132	V/A	4/12	rs201533100	0.2% (1/498
T>T/A	505	106	36	T/S	2/12	rs147562047	0.2% (1/498
G>G/T	770	371	124	T/N	3/12		0.2% (1/498
G>G/T	1137	738	246	S/R	6/12		0.4% (2/498)

Table 4. *PKM* missense variants detected in 11 clinical trial participants .

Examination of *PKLR* genetic variants detected in seven clinical trial participants. TFV-MP phosphorylation by the kinase *PKLR* been demonstrated in PBMC and vaginal tissue [5]. For the DNA reference sequence corresponding to *PKLR* (NM_000298.5), five variants comprised of SNVs and deletions that are predicted to be reflected in the amino acid sequence were detected across seven individuals. These variants are shown in Table 5, with further information available in Supplementary Table 4. Two previously unreported variants were predicted to have a deleterious impact on protein function. The amino acid substitution of tryptophan for glycine at the protein position 406 is predicted to be deleterious as is the previously unreported N133K mutation at the protein level. Of the five variants detected in this population, three had been previously reported and are in the dbSNP. The variant rs147659527 (NM_000298.5:c.814C>G) was detected in two individuals as was the variant rs61755431 (NM_000298.5:c.1706G>A). The variant rs116100695 (NM_000298.5:c.1456C>T) was detected in one individual. Of all the individuals carrying variants in the *PKLR* gene, all 7 individuals are predicted to be carrying deleterious variants. This results in a 1.4% frequency (7/498) of individuals exhibiting deleterious variants.

Variant (ref.>alt.)	cDNA Position	CDS Position	Protein Position	Amino Acids	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
C>C/A	1255	1216	406	G/W	8/11	deleterious(0)	probably_damaging(1)		0.2% (1/498)
G>G/C	853	814	272	L/V	6/11	deleterious(0)	probably_damaging(0.974)	rs147659527	0.2% (1/498)
C>C/T	1745	1706	569	R/Q	11/11	deleterious(0.04)	benign(0.399)	rs61755431	0.4% (2/498)
G>G/C	853	814	272	L/V	6/11	deleterious(0)	probably_damaging(0.974)	rs147659527	0.2% (1/498)
G>G/T	438	399	133	N/K	4/11	deleterious(0.03)	probably_damaging(0.999)		0.2% (1/498)
G>G/A	1495	1456	486	R/W	10/11	deleterious(0)	probably_damaging(0.996)	rs116100695	0.2% (1/498)

Table 5. *PKLR* missense variants detected in seven clinical trial participants.

Analysis of individuals carrying genetic variants in candidate emtricitabine-activating kinases *CMPK1, DCK, PGK1,* and *TK1*. FTC is an NRTI cytidine analog that must be phosphorylated intracellularly to FTC-MP, FTC-DP, and finally FTC-TP to form the active triphosphate that inhibits HIV reverse transcriptase. We investigated the presence of genetic variants predicted to be present at the amino acid level in the kinases *CMPK1, DCK, PGK1,* and *TK1* in 498 clinical trial participants across 13 clinical research sites in the United States. Of these 498 individuals, 34 exhibited variants in *CMPK1*, 11 exhibited variants in *DCK*, and 2 individuals displayed variants for the *PGK1* gene. The 33 individuals with detected variants in *CMPK1* indicate a 6.8% (34/498) frequency of individuals carrying variants in this population. For the *DCK* gene, 12 individuals carrying variants yielded a 2.4% (12/498) frequency. In *PGK1*, two of 498 individuals exhibited variants in this gene (0.4% frequency, 2/498 individuals).The kinase *TK1* did not have any individuals with variants detected to affect the amino acid level of this protein. Distribution of variants among individuals is depicted in a Venn diagram in Figure 3. Within this population, three clinical trial participants exhibited variants in more than one kinase. *CMPK1* and *DCK* variants were detected in 2 individuals, while one individual exhibited variants in all three kinases with detected variants at the amino acid level.

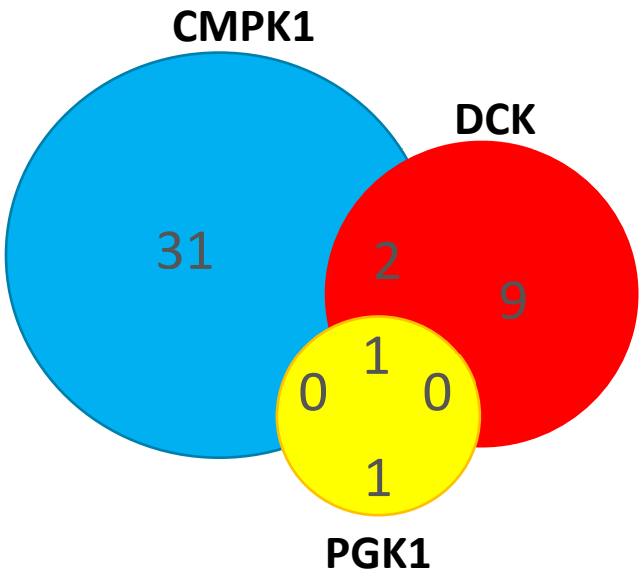

Figure 3. Distribution of individuals carrying genetic variants at the amino acid level in candidate kinases that phosphorylate FTC. Each circle represents a nucleotide kinase that was sequenced, with *CMPK1* in blue, *DCK* in red, and *PGK1* in yellow. Numerical values indicate the number of individuals detected to carry a single nucleotide variation or deletion. Overlapping regions of each circle indicate the number of individuals with genetic variants in more than one kinase.

Investigation of *DCK* genetic variants observed in 12 clinical trial participants. *DCK* has been reported to show activity toward deoxycytidine analogs (thymidine kinase and deoxycytidine kinase in HIV-infected children) [7] [8], rendering this kinase a prospective activator of FTC, a cytidine analog. Given the reported activity of *DCK* on deoxycytidine analog, one can envision possible activity of this kinase toward the cytidine analog FTC. In the 498 individuals sequenced for the kinase *DCK* (NM_000788.2), 12 individuals were detected to carry five unique genetic variants predicted to affect translation at the amino acid level of this gene. All individuals carrying variants exhibited heterozygosity for these SNVs. The variants detected are shown in Table 6, with further detail found in Supplementary Table 5. Of these six genetic variants, two variants were previously unreported, with one variant with the protein mutation S93F predicted to be deleterious and the other variant resulting in a stop codon. The three previously reported variants are accessible in the SNP database and are listed as follows:

rs67437265 (NM_000788.2:c.364C>T) was detected in six individuals, rs66878317 (NM_000788.2:c.70A>G) was detected in one individual, and the SNP rs144479260 (NM_000788.2:c.261G>A) was detected in one individual. Across all individuals carrying variants for *DCK,* only one individual presented a variant predicted to be deleterious, resulting in a 0.2% (1/498) frequency.

Variant (ref.>alt.)	cDNA Position	CDS Position	Protein Position	Amino Acids	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
C>C/T	652	364	122	P/S	3/7	tolerated(0.06)	probably_damaging(1)	rs67437265	0.4% (2/498)
C>C/T	566	278	93	S/F	3/7	deleterious(0)	probably_damaging(0.969)		0.2% (1/498)
A>A/G	358	70	24	I/V	1/7	tolerated(0.65)	benign(0.002)	rs66878317	0.2% (1/498)
C>C/T	652	364	122	P/S	3/7	tolerated(0.06)	probably_damaging(1)	rs67437265	0.8% (4/498)
G>G/T	739	451	151	E/*	4/7				0.6% (3/498)
G>G/T	549	261	87	E/D	3/7	tolerated(0.61)	benign(0)	rs144479260	0.2% (1/498)

Table 6. *DCK* missense variants detected in clinical trial participants across clinical research sites located in the United States.

***TK1* genetic variants detected in five clinical trial participants did not result in amino acid changes.** In this study, no exonic, or missense variants for *TK1* (NM_003258.4) were detected in any clinical trial participants. However, a previously reported intron variant rs2661681 (NM_003258.4:c.394-4G>A) was detected in five individuals, yielding a 1% (5/498) frequency for the only variant detected for this kinase in this study. Because this variant is in the intron, there was no functional impact prediction available.

Examination of *CMPK1* genetic variants detected in 34 clinical trial participants. *CMPK1* has been shown to phosphorylate 3TC-MP, a monophosphorylated cytidine analog to 3TC-DP [9]. Similarities in 3TC and FTC structure indicate *CMPK1* as a kinase with prospective activity toward phosphorylating FTC-MP to FTC-DP. Using the reference DNA sequence NM_016308.2 for *CMPK1*, seven variants across 34 individuals were detected. All individuals were heterozygous for the variants they exhibited. These variants are shown in Table 7 and with further detail in Supplementary T able 8. Of these seven variants, three variants were previously

unreported. An unreported variant translating to an arginine to serine mutation at the 74 protein position was predicted to be deleterious, with one individual exhibiting this variant. A second unreported variant was also predicted to be deleterious, with an amino acid substitution from arginine to isoleucine at the 183 protein position. One individual in this study was observed to carry this variant. The third unreported *CMPK1* missense variant was detected in five individuals and is predicted to affect the 192 protein position, substituting an asparagine residue for a lysine residue. This variant was predicted to be tolerated. The remaining four variants detected in individuals for this kinase were listed in the SNP database. The variant rs80046781 (NM_016308.2:c.196G>T) was detected in 9 individuals and is predicted to be deleterious, yielding a 1.8% (9/498) frequency of this variant. A second previously reported variant, rs35687416 (NM_016308.2:c.240G>T) was also predicted to be deleterious, with 11 individuals exhibiting this variant (2.2% frequency, 11/498). The variant rs139782187 (NM_016308.2:c.498G>T) was detected in one individual and was predicted to be deleterious. Finally, the fourth variant rs72553947 (NM_016308.2:c.344A>G) was predicted to be tolerated and was found in four individuals. Overall, the frequency of individuals exhibiting deleterious variants for *CMPK1* was 5.0% (25/498).

Variant (ref.>alt.)	cDNA Position	CDS Position	Protein Position	Amino Acids	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
G>G/T	345	196	66	A/S	2/6	deleterious(0)	possibly_damaging(0.901)	rs80046781	(9/498)
G>G/T	371	222	74	R/S	2/6	deleterious(0.03)	benign(0.374)		(2/498)
G>G/T	697	548	183	R/I	4/6	deleterious(0)	probably_damaging(0.998)		(1/498)
G>G/T	389	240	80	Q/H	2/6	deleterious(0)	benign(0.31)	rs35687416	(11/498)
G>G/A	468	319	107	E/K	3/6	deleterious(0.01)	benign(0.023)	rs72553946	(1/498)
G>G/T	647	498	166	R/S	4/6	deleterious(0)	probably_damaging(0.999)	rs139782187	(1/498)
A>A/G	493	344	115	N/S	3/6	tolerated(0.84)	benign(0.001)	rs72553947	(4/498)
G>G/T	725	576	192	K/N	5/6	tolerated(0.07)	benign(0.147)		(5/498)

Table 7. *CMPK1* missense variants detected in 34 clinical trial participants.

***PGK1* genetic variants detected in two clinical trial participants.** L-Fd4C-DP, a cytidine diphosphate analog like FTC-DP has been shown to be phosphorylated by *PGK1* to form 3TC-TP, a nucleotide triphosphate that competitively inhibits HIV reverse transcriptase [10]. We detected one previously unreported variant and one previously reported variant in *PGK1*. Neither variant was predicted to be deleterious.

Variant	Genotype	cDNA Position	CDS Position	Protein Position	Amino Acids	Exon	Sift	PolyPhen	dbSNP ID	Variant Frequency
G>G/T	het	605	433	145	A/S	5/11	tolerated(0.16)	benign(0.023)		0.2% (1/498)
G>A/A	hom	355	183	61	M/I	3/11	tolerated(0.1)	benign(0.062)	rs11541568	0.2% (1/498)

Table 8. *PGK1* missense variants detected in two clinical trial participants.

Discussion

This study identifies and proposes genetic variants in candidate kinases for FTC activation, hypothesizing DCK and TK1 to phosphorylate FTC to FTC-MP, CMPK1 to phosphorylate FTC-MP to FTC-DP, and PGK1 to phosphorylate FTC-DP to FTC-TP, the pharmacologically active anabolite of FTC. In this work, we detected six previously unreported variants across these kinases. Further, we contribute more unreported variants to the growing knowledge of genetic variation in kinases previously identified to phosphorylate TFV. In the

genes that encode the TFV-activating kinases AK2, CKM, PKM, and PKLR, we detected 11 previously unreported variants across these kinases. In the kinases DCK, TK1, CMPK1, and PGK1, the enzymes proposed to activate FTC, we detected six previously unreported variants. The analysis implemented in this study conformed to the American College of Medical Genetics and Genomics clinical laboratory standards for next-generation sequencing in order to increase confidence in the variants we detected [11]. This analysis has been previously described [5].

Focusing on AK2, the kinase we have previously demonstrated to phosphorylate TFV to TFV-MP, displayed three previously reported genetic variants upon sequencing for this gene. The variant rs138577419 was detected in one individual in this study, yielding a frequency of 0.2% (1/498). Comparing this frequency to the overall frequency listed in the Exome Aggregation Consortium (ExAC), a large-scale aggregation of human genomic data including over 60,000 individuals [12], ExAC reports a frequency of 0.2%. Four of the seven subpopulations reported frequencies, with the non-Finnish European reporting 0.4%, African reporting 0.1%, African reporting 0.02%, and the sub-population Other reporting 0.5% frequency of this variant. A second previously reported variant, rs12116440 was detected in two individuals in this study, exhibiting a 0.4% frequency. This frequency is lower than the overall reported frequency for this variant reported in the ExAC database, which was 0.7%. This variant was detected in all sub-populations except for the East Asian subpopulation, which showed no frequency for this variant. This variant was also detected in one individual in a previous genotyping study carried out by our lab [5]. The third variant detected for AK2 that has been previously reported, rs61750965 also displayed a 0.4% (2/498) frequency in this study. A previous study sequencing this kinase carried out by our lab also detected this variant in one individual [5]. This variant is reported at an overall frequency of 0.4% in the ExAC database, with all sub-populations reporting frequencies for this variant. Of the subpopulation frequencies reported for this variant, the highest reported is in the South Asian

population, with 0.7% and the lowest reported is in the East Asian population, with a 0.01% frequency.

In DCK, one of two candidate kinases hypothesized to phosphorylate FTC to FTC-MP, DCK, three variants detected in our study have been previously reported in the SNP database. The variant rs67437265 was detected in six individuals in this study, with a frequency of 1.2% (6/498). Comparing this frequency to data reported by ExAC, we find that the overall population frequency for this variant is 1.9%. This includes African, East and South Asian, Latino, and European populations. Another previously reported variant, rs66878317 was detected in one individual, yielding a 0.2% frequency in this study. The frequency for this variant reported in ExAC was also reported at 0.2% across all populations, with East Asian and Finnish European populations included in the ExAC database reporting 0% frequency and African populations reporting 2.8% frequency. The last previously reported variant rs144479260 was reported in the ExAC database at 0.015% frequency, with only non-Finnish European (0.02% frequency), Finnish European (0.015% frequency), and Latino (0.017% frequency) populations reporting frequencies for this variant. These frequencies are lower than what was detected for this variant in this study, with one individual (0.2%, 1/498) exhibiting this variant. Sequencing this population for missense variants in DCK identified one previously undetected variant that was predicted to be deleterious. This variant translates to a mutation at the 93 protein position, reported to be located in an alpha helix and four amino acids away from a substrate binding site at the 97th amino acid [13]. The substitution from serine to phenylalanine reported at this position (S93F), changes the residue from a polar, uncharged, and relatively small side chain on the serine to a large, hydrophobic benzene ring on the resulting phenylalanine. This substitution coupled with the proximity to the reported substrate binding site could impact kinase function.

Sequencing for the gene encoding CMPK1, four variants detected were found to be reported in the SNP database. The variant rs80046781 was detected in 9 individuals, however no reported frequency was available in the ExAC database. Predicted to result in a deleterious functional impact for this kinase, the genetic variant rs80046781 translates to the amino acid substitution at the 66 protein position A66S. This position is three amino acids away from a reported nucleotide monophosphate binding domain [14] and the amino acid substitution changes from a hydrophobic side chain in alanine to a polar side chain in serine. *In vitro* studies using this purified enzyme have been previously carried out investigating CMPK1 activity on another deoxycytidine analog, 2',2'-difluorodeoxycytidine and have demonstrated activity kinase activity toward the monophosphorylated form of this compound [14]. Taken together, the nucleotide binding domain location, amino acid substitution, and demonstrated nucleotide analog phosphorylation information available for this kinase and the genetic variant detected could provide an explanation for why this variant is predicted to negatively impact protein function. A second previously reported variant rs35687416 was also predicted to be deleterious, with 11 individuals exhibiting this variant (2.2% frequency, 11/498). Comparing this frequency to the frequency of this variant as reported in the ExAC database, all sub-populations included in the database exhibit frequency for this variant, resulting in an overall 4.5% frequency. The two sub-populations displaying the highest frequency for this variant were East Asian, with 12% frequency and Finnish European with 6.7% frequency. The variant rs72553947 was detected in four individuals, resulting in a 0.8% frequency (4/498). This frequency is close to the overall frequency reported in the ExAC database, which reported a frequency across all populations at 0.6%. All populations reported frequencies for this variant, with the population reporting the highest frequency at 3.1% being the East Asian population. The fourth previously reported variant for CMPK1 was detected in one individual in our study and was reported in the ExAC database at a frequency of less than 0.001%. Only one subpopulation included in the database

reported frequencies for this variant, with the African subpopulation presenting a 0.001% frequency for this reported variant.

The candidate kinase for phosphorylation of FTC-DP to FTC-TP, PGK1, exhibited one genetic variant rs11541568 included in the SNP database. This variant was detected in one individual and reported at an overall frequency of 0.001% in the ExAC database. The African subpopulation included in the database was the only one to report frequency of this variant, showing a frequency of 0.1% within that sub-population. This variant was not predicted to be deleterious, likely because of its location. Resulting in an amino acid mutation of M61I, this residue is not near any catalytic sites or nucleotide binding sites [15].

In sequencing the kinases DCK, TK1, CMPK1, and PGK1, we are suggesting a pathway for emtricitabine activation. The kinases in this study predicted to phosphorylate emtricitabine need to be validated for activity *in vitro* toward emtricitabine and its phosphorylated anabolites, despite previously demonstrated activity toward other cytidine analogs. The predicted functional impact of the identified variants in these kinases also ultimately requires confirmation. Purified protein assays using site-directed mutagenesis would be beneficial toward confirmation of this predicted diminished activity.

The variants we have identified for TFV activation build further upon the concept we have previously put forth that protection against HIV infection using PrEP may yield inter-individual results. Observation of previously unreported genetic variants in the kinases AK2, CKM, PKM, and PKLR contribute to the notion that there could be a genetic basis for variation in TFV phosphorylation and therefore protection from HIV. Functional deleterious predictions of three previously unreported genetic variants in CKM and two previously unreported genetic variants in PKLR bolster this concept. Further, the pathway of emtricitabine activation and the

resulting variants from the kinases we have proposed adds another element that could contribute to the variation in efficacy of PrEP despite adherence. Continuing to build upon the knowledge of variants in the kinases shown to activate TFV as well as addressing FTC phosphorylation encompasses all components of current FDA-approved PrEP formulations, allowing for a fuller understanding of the discrepancies we see in HIV protection.

References

[1] US Food and Drug Administration, "FDA Approves First Medication to Reduce HIV Risk.," FDA Consumer Health Information, 2012.

[2] Grant, R.M., Lama, J.R., Anderson, P.L., McMahan, V., Liu, A.Y., Vargas, L., et al., 2010. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N. Engl. J. Med. 363 (27), 2587–2599.

[3] Thigpen, M.C., Kebaabetswe, P.M., Paxton, L.A., Smith, D.K., Rose, C.E., Segolodi, T.M., et al. 2012. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. N. Engl. J. Med. 367 (5) 423-434.

[4] Baeten, J.M., Donnell, D., Ndase, P., Mugo, N.R., Campbell, J.D., Wangisi, J., et al., 2012. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N. Engl. J. Med. 367 (5), 399–410.

[5] Lade JM, To EE, Hendrix CW, Bumpus NN. "Discovery of Genetic Variants of the Kinases That Activate Tenofovir in a Compartment-specific Manner," *EBioMedicine,* vol. 2, no. 9, p. 1145–1152, 2015.

[6] Dando T, Wagstaff A, "Emtricitabine/Tenofovir Disoproxil Fumarate.," *Drugs,* pp. 2075-2082, 2004.

[7] Furman PA, Davis M, Liotta DC, Paff M, Frick LW, Nelson DJ, Dornsife RE, Wurster JA, Wilson LJ, Fyfe JA, et al. "The Anti-Hepatitis B Virus Activities, Cytotoxicities, and Anabolic Profiles of the (-) and (+) Enantiomers of cis-5-Fluoro-1-[2-(Hydroxymethyl)-1 ,3-Oxathiolan-5-yl]Cytosine.," *Antimicrobial Agents and Chemotherapy.,* vol. 36, no. 12, pp. 2696-2692, 1992.

[8] Balzarini J, Herdewijn P, De Clercq E., "Differential patterns of intracellular metabolism of 2',3'-didehydro-2',3'-dideoxythymidine and 3'-azido-2',3'-dideoxythymidine, two potent anti-human immunodeficiency virus compounds.," *J Biol Chem. ,* vol. 264, no. 11, pp. 6127-6133, 1989.

[9] Pasti C, Gallois-Montbrun S, Munier-Lehmann H, Veron M, Gilles AM, Deville-Bonne D., "Reaction of human UMP-CMP kinase with natural and analog substrates," *Eur. J. Biochem. ,* vol. 270, p. 1784–1790, 2003.

[10] Krishnan P, Fu Q, Lam W, Liou JY, Dutschman G, Cheng YC., " Phosphorylation of pyrimidine deoxynucleoside analog diphosphates: selective phosphorylation of L-nucleoside analog diphosphates by 3-phosphoglycerate kinase." *The Journal of Biological Chemistry. ,* vol. 277, no. 7, pp. 5453-5459, 2002.

[11] Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., et al., 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med*. 15 (9), 733–747.

[12] Exome Aggregation Consortium, et al. 2015. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv.* Cold Spring Harbor Laboratories.

[13] Sabini E, Hazra S, Ort S, Konrad M, Lavie A., "Structural Basis for Substrate Promiscuity of dCK," *Journal of Molecular Biology,* vol. 378, no. 3, pp. 607-621, 2008.

[14] Van Rompay AR, Johansson M, Karlsson A., "Phosphorylation of Deoxycytidine Analog Monophosphates by UMP-CMP Kinase: Molecular Characterization of the Human Enzyme.," *Molecular Pharmacology,* vol. 56, no. 3, pp. 562-569, 1999.

[15] Michelson AM, Blake CC, Evans ST, Orkin SH., "Structure of the human phosphoglycerate kinase gene and intron-mediated evolution and dispersal of the nucleotide-binding domain.," *Proc. Nati. Acad. Sci. USA,* vol. 82, pp. 6965-6969, 1985.

Chapter 4: Imaging tenofovir and tenofovir diphosphate using matrix assisted laser desorption ionization mass spectrometry imaging.

Abstract

A method to image tenofovir and its phosphorylated anabolite tenofovir diphosphate using matrix assisted laser desorption/ionization coupled to mass spectrometry has been developed. To detect tenofovir and tenofovir diphosphate, four different matrices were tested in both positive and negative ionization modes. Moving forward with the matrices generating the best sensitivity for each compound, α-hydroxycinnamic acid was used to detect tenofovir and 9-aminoacridine was used to detect tenofovir diphosphate. TFV was detected at the highest spotted concentration of 10 μM and the lowest detectable spotted concentration of 125 nM. For TFV-DP detection, the highest spotted concentration was 10 μM and the lowest detectable spotted concentration at 78 nM. This novel method can be applied to image tenofovir in tissue and detect tenofovir diphosphate using MALDI-MS.

Introduction

Along with being approved for use in HIV treatment, the nucleotide reverse transcriptase inhibitor tenofovir (TFV) is approved for and continues to be explored as a pharmacological approach toward protection from HIV infection [1-3]. Known as HIV pre-exposure prophylaxis, or PrEP, this strategy involves an HIV-uninfected individual taking TFV to prevent HIV infection. Currently, the only FDA approved route of administration of TFV for PrEP is oral [3], but topical routes of administration are under active investigation in HIV-susceptible tissues [4, 5, 6].

In order to become pharmacologically active, TFV must be phosphorylated intracellularly. A prior study done by our laboratory has shown TFV to be phosphorylated to tenofovir-monophosphate (TFV-MP) by adenylate kinase 2 in peripheral blood mononuclear cells

(PBMC), vaginal, and colorectal tissue [7]. The active TFV anabolite, however, is tenofovir-diphosphate (TFV-DP). A nucleotide triphosphate analog, TFV-DP competitively inhibits HIV reverse transcriptase and prevents viral DNA replication [1]. TFV-DP is formed by phosphorylation of TFV-MP by creatine kinase, muscle in colorectal tissue [7]. In PBMC and vaginal tissue, TFV-MP has been shown to be phosphorylated by pyruvate kinase, liver and red blood cell and pyruvate kinase, muscle to form TFV-DP [7].

Taken together, the compartment specificity of TFV activation demonstrated in our previous work and investigations of the efficacy of topical PrEP formulations call for a deeper understanding of TFV tissue pharmacokinetics. TFV and TFV-DP concentrations have been examined in compartments vulnerable to HIV infection such as rectal, vaginal, and cervical tissues [6, 8, 9, 10]. While these studies have been helpful in confirming and measuring the amount of drug able to penetrate HIV-susceptible tissues, quantifications of TFV and TFV-DP are measured using liquid chromatography coupled to tandem mass spectrometry (LC/MS) from tissue homogenates [6, 9]. Measurements from homogenates are only able to provide a quantification of drug that is essentially the amount found throughout the tissue. Using these methods to examine TFV and TFV-DP in tissues, any drug distribution information is lost when the tissue is homogenized.

Along with LC/MS methods, another way drug is measured in tissue is using whole body autoradiography [11]. Currently the standard technique used in preclinical studies to determine drug distribution [12], this technique this technique uses radioactivity incorporated into a molecule of interest. This is then administered in an animal model and the resulting emissions are detected and imaged. Despite the ability to provide spatial distribution information, however, this technique is limited in that it is only able to detect the distribution of the radioactive compound. Because of this, it is possible that metabolites of the drug will not be detected.

The two-dimensional nature in which data are collected using matrix assisted laser desorption/ionization coupled to imaging mass spectrometry is able to yield drug distribution information in tissue [13]. Like LC/MS methods, detection of compounds and any metabolites is possible due to analysis using the mass spectrometer. Similar to inquiry using whole body autoradiography spatial distribution information is attainable due to the movement of the sample plate in MALDI-IMS. In this study, we harness this technology and develop a method to detect TFV and its phosphorylated anabolite TFV-DP using a MALDI-Orbitrap XL mass spectrometer [14].

TFV has previously reported to be detected using MALDI technology in a IR-laser matrix assisted laser desorption electrospray ionization (IR-MALDESI) [15]. Using this technique, distribution of TFV in cervical tissues was investigated. More conventional uses of MALDI have also been applied to detect TFV. TFV has been detected using MALDI coupled to a triple quadrupole mass spectrometer in plasma from HIV-infected adults [16]. This study develops an assay for the detection of TFV and TFV-DP using MALDI coupled to an LTQ Orbitrap XL mass spectrometer. Not only are we able to detect the parent drug TFV, but we also report detection of its phosphorylated active metabolite TFV-DP using MALDI. This technique can be applied to gain spatial distribution information of these anti-HIV drugs and inform future clinical studies investigating PrEP efficacy.

Experimental/Materials and methods

Materials. Solvents for the drug standards and matrix solutions were Optima™ uHPLC grade water, methanol, acetonitrile, and ethanol (200 proof) from Decon™ Labs were all obtained from Fisher Scientific (Waltham, MA). Trifluoroacetic acid was purchased from Sigma-Aldrich (St. Louis, MO). The matrices tested were obtained as follows: α-cyano-4-hydroxycinnamic acid (CHCA), 9-aminoacridine (9AA), anthranilic acid (AA), nicotinic acid (NA), and 3-hydroxypicolinic acid (3HPA) were all procured from Sigma-Aldrich (St. Louis, MO). The matrix compound 2,5-dihydroxybenzoic acid (DHB), 99% from ACROS Organics™ was purchased from Fisher Scientific (Waltham, MA). Tenofovir was obtained from the NIH AIDS Reagent Program, Division of AIDS (DAIDS), NIAID, NIH. Tenofovir diphosphate was purchased as a triethylamine salt from Toronto Research Chemicals (North York, ON, Canada).

Drug standard preparation. TFV and TFV-DP were diluted separately using uHPLC grade water to 10 mM as a starting concentration and further diluted for MALDI-MS analysis. Solutions were stored at -20° C.

Matrix solution preparation. Matrix solutions that were investigated for their suitability to detect TFV or TFV-DP are detailed in Table 1. Following preparation of matrix solutions, all solutions were sonicated in a water bath for 30 minutes prior to application to the drug standards for analysis.

Matrix Used	Concentration (mg/mL)	Solvent
α-cyano-4-hydroxycinnamic acid	10	70% acetonitrile, 30% water, 0.3% trifluoroacetic acid
9-aminoacridine	10	90% ethanol, 10% water
anthranilic acid, nicotinic acid, diammonium citrate	6.17, 5.54, 12.44	55% water, 45% acetonitrile
3-hydroxypicolinic acid	40	70% acetonitrile, 30% water, 0.3% trifluoroacetic acid
2,5-dihydroxybenzoic acid	40	70% acetonitrile, 30% water, 0.3% trifluoroacetic acid

Table 1. Matrix solutions investigated for suitability to detect TFV or TFV-DP.

Sample preparation. To determine optimal settings for the detection of TFV and TFV-DP, drug standards were mixed with matrix solution at a 1:1 ratio and pipetted onto a 384 spot steel target plate. Optimal mass spectrometer settings to detect TFV were determined using a 10 mg/mL concentration of CHCA and a final TFV concentration of 10 μM. TFV-DP detection was optimized using a 10 mg/mL concentration of 9-AA matrix mixed with TFV-DP standard to 10 μM. To establish a method for imaging TFV and TFV-DP on tissue, 1.5 μL dilutions of each drug standard were pipetted onto the steel target plate. Serial dilutions at 10, 5, 2.5, 1.25, 0.625, 0.3125, 0.156, and 0.78 μM of TFV and TFV-DP were spotted at a volume of 1.5 μL. To ensure specific detection of each compound, 1.5 μL of water (as vehicle) was also pipetted and analyzed for the presence of TFV or TFV-DP.

LTQ Orbitrap XL analysis. A MALDI-MS/MS method was developed and optimized to detect TFV and TFV-DP using a MALDI source coupled to a LTQ Orbitrap XL mass spectrometer (Thermo Scientific, Pittsburgh, PA). Laser energy for ionization of TFV and TFV-DP in the MALDI source was used at 20 μJ. Automatic gain control target settings to use the Orbitrap mass analyzer (FTMS) were set at 1 X 10^6 ions for a full MS scan, 5 X 10^4 ions for a selected ion monitoring (SIM) scan, and 2 X 10^5 ions for MS/MS analysis. The following parameters were specified on the FTMS mass analyzer for TFV detection: positive ionization mode, mass

resolution at 100,000. Normalized collision energy (NCE) using collision-induced dissociation for generation of TFV fragment ions was tested at 15%, 20%, and 35%. To identify TFV, the transition 288.08560→176 (176.09191 ± 5 ppm)was monitored. TFV-DP detection was achieved using the following parameters specified: negative ionization mode, mass resolution at 100,000. To generate fragment ions for TFV-DP identification, collision-induced dissociation was performed. NCE at 10%, 15%, and 20% were tested. Identification of TFV-DP was determined by monitoring the transition 446.00372→348 (348.0975 ± 5 ppm). Spotted drug standards were imaged with the camera on the MALDI source using Thermo Tune Plus software (Thermo Scientific, Pittsburgh, PA). Raster size of each laser step was used at 400 μm step size.

Data processing and analysis. ImageQuest software (Thermo Scientific, Pittsburgh, PA) was used to visualize the fragment ions generated from TFV or TFV-DP detection. To determine the spatial distribution of TFV, the mass to charge ratio at 176.09296 was extracted from the spectra collected using a narrow mass tolerance window of 5 ppm. TFV-DP distribution was localized using the mass to charge ratio at 348.0975 with the same mass range window at 5 ppm.

Results

TFV detection is achieved using CHCA matrix while TFV-DP is ionized using 9AA. Several matrices were tested for their suitability to detect either TFV or TFV-DP. Of all the matrices tested, only CHCA was able to ionize TFV at the starting concentration of the drug standard we tested at 10 μM. Using the CHCA matrix, TFV was ionized in positive ion mode, yielding a protonated parent ion m/z ratio at 288.08560 ± 5 ppm. To detect TFV-DP, all four matrices were also tested. The matrix 9AA was the only matrix that allowed for detection of TFV-DP at 10 μM. The 9AA matrix ionized TFV-DP in negative ion mode, generating a parent ion at m/z = 446.00372 ± 5 ppm.

Normalized collision energy of 35% provides optimal fragmentation for TFV detection with CHCA matrix using the LTQ Orbitrap XL. Using the matrix CHCA to ionize TFV, generation of fragment ions of TFV was investigated by performing collision-induced dissociation at incrementally higher NCE (Figure 1). At 15% NCE, no fragmentation of TFV (m/z = 288.08378) was observed. Increasing NCE to 20% yielded fragmentation of TFV, producing fragment ions at m/z = 176.09188 and m/z = 206.10225. Water loss of TFV at m/z = 270.07345 was also observed. NCE at 35% generated the highest abundance of the fragment ion at m/z = 176.09191 from the parent TFV ion at m/z = 288.03838.

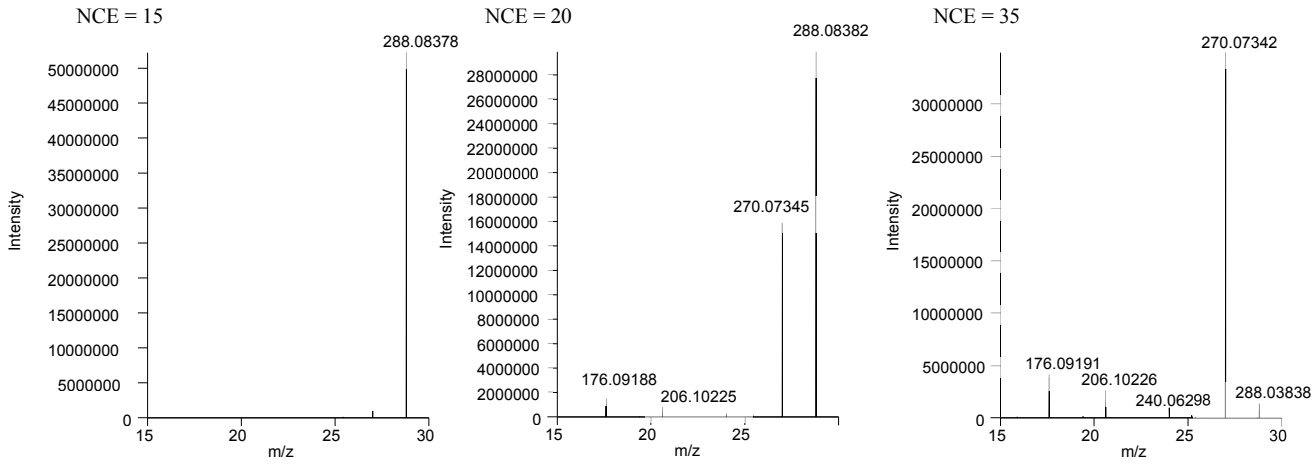

Figure 1. Total ion chromatogram fragmentation spectra of TFV at increasing normalized collision energies (NCE). Collision-induced dissociation was used to generate fragment ions to identify TFV. NCE at 15%, 20%, and 35% were tested. NCE at 15% shows no fragmentation of TFV. The NCE at 35% yielded the greatest intensity for the fragment ion of TFV using the transition at m/z = 288.08560→176.

Spatial distribution of TFV on a steel target plate can be determined using MALDI-MS imaging. Following optimization of TFV detection using CHCA matrix, detection methods were applied toward investigating TFV localization on a steel target plate. TFV standard at 10 µM final concentration mixed in a 1:1 ratio with CHCA matrix can be visualized when using the extracted fragment ion at m/z = 176 (Figure 2). All spectra included in the images in the middle panel of figure 2 contain detectable peaks at m/z = 176.09296 with a 5 ppm mass tolerance window. To ensure specific identification of TFV, water was combined in a 1:1 ratio with CHCA matrix solution. No spectra collected in the water sample had a peak at m/z = 176.09296 ± 5 ppm. On a steel target plate, TFV was detected at the highest spotted concentration of 10 µM and the lowest detectable spotted concentration of 125 nM.

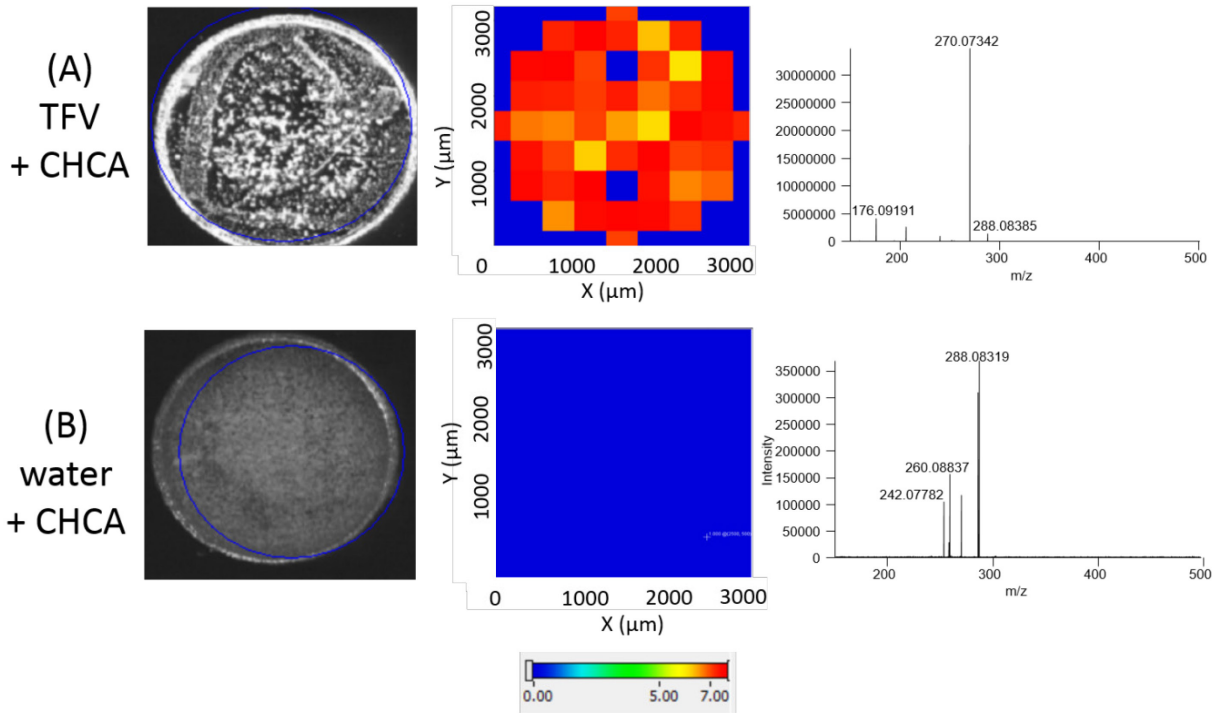

Figure 2. Distribution of TFV standard on a steel target plate using extracted ion MS/MS data. (A) TFV drug standard solution is mixed 1:1 with CHCA matrix solution, with 1.5 µL of the resulting solution pipetted onto a steel target plate. The resulting molecular image shows localization of TFV using the fragment ion at m/z = 176.09296 ± 5 ppm. The spectrum displayed is an average spectra across the entire sample. (B) Water is mixed in a 1:1 ratio with CHCA matrix solution. A volume of 1.5 µL of the resulting mixture is pipetted onto a steel target plate for MALDI-IMS analysis. When m/z = 176.09296 ± 5 ppm is extracted from the resulting spectra of this dataset, this peak is absent from all spectra collected.

Differences in TFV abundance can be detected using MALDI-MS imaging. This technique can demonstrate differences in abundance of tenofovir when a drug standard is pipetted onto a glass slide. Following application with CHCA matrix as described above, drug standards at 10 µM, 5 µM, and 2.5 µM were detected at differing intensities. The fragment ion peak at 288.0856 → 176.09296 ± 5 ppm was detected at 933.9 intensity for the 2.5 µM drug standard. At 5 µM and 10 µM, the fragment ion peak was detected at 2839 and 6153 respectively.

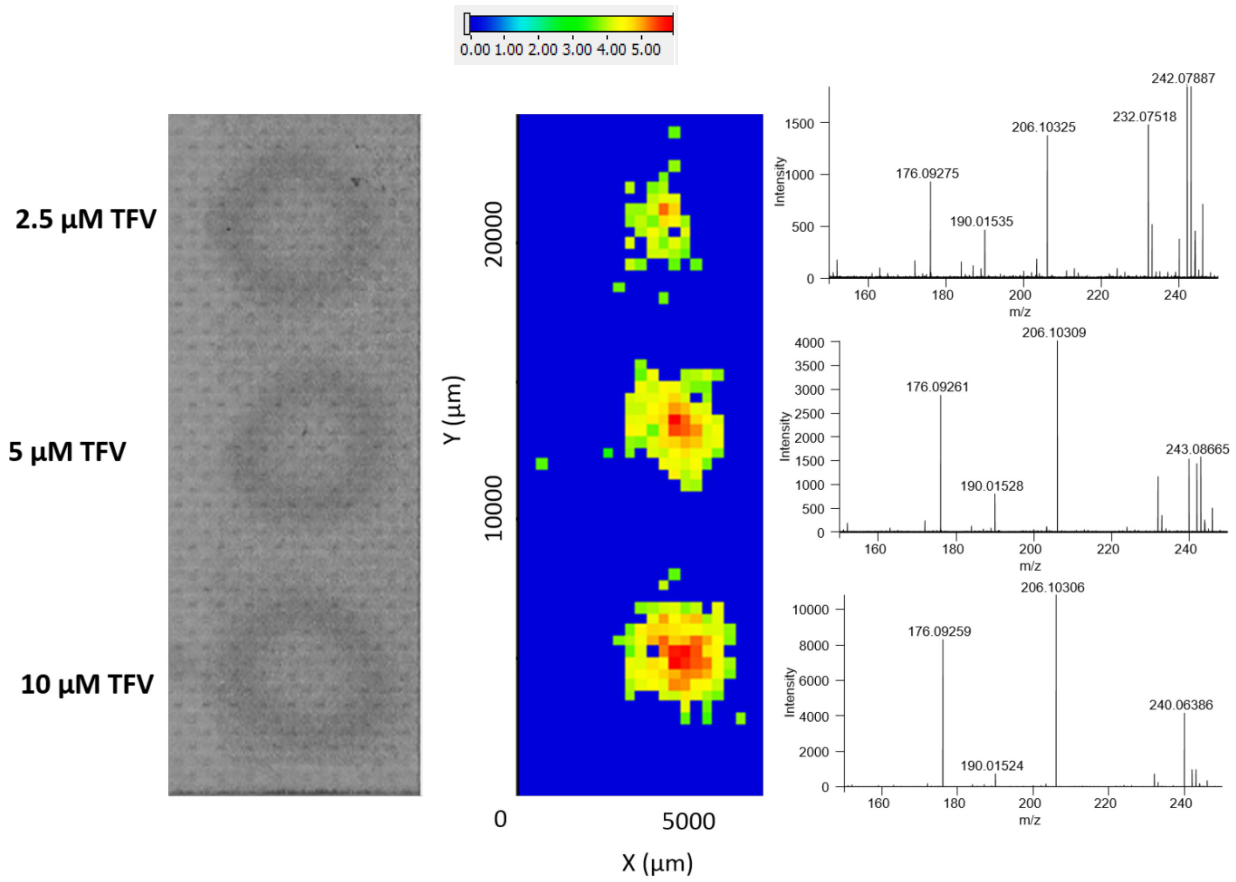

Figure 3. Detection of TFV drug standard at different concentrations shows differences in intensity. Depicted in the figure above is an optical image of the three areas in which the drug standard was spotted and subsequently coated with CHCA matrix. The middle panel shows a molecular image of all collected spectra from which the fragment ion 288.08560→176.06296 was detected. Finally, the fragmentation spectra in the rightmost panel show increasing intensity of the peak at 176.09296 ± 5 ppm at increasing concentrations of TFV drug standard are imaged.

Detection of TFV in cervical biopsy. TFV gel (40 mg of TFV) in water with edetate disodium, citric acid, glycerin, methylparaben, propylparaben, and hydroxyethycellulose was applied topically to vaginal tissue of a clinical trial participant. After application, five hours were allowed to elapse and a cervical biopsy was collected. We are able to detect TFV in the biopsy and observe distinct heterogeneity of TFV detection throughout the biopsy.

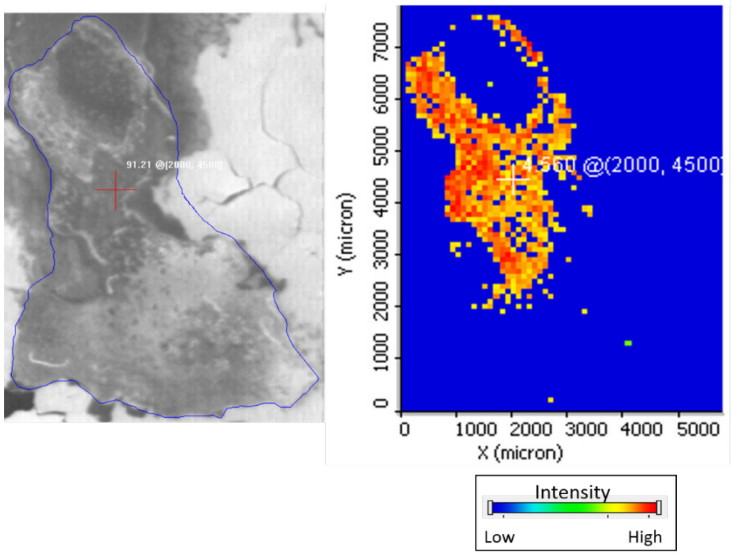

Figure 4. TFV distribution in a cervical biopsy is detected in a heterogeneous manner. TFV is not detected throughout the biopsy in a homogenous manner. Instead, we observe TFV localization in one area of the biopsy but not its entirety.

Normalized collision energy of 20% provides optimal fragmentation for TFV-DP detection with 9AA matrix using the LTQ Orbitrap XL. Subsequent to determining the suitability of 9AA as an effective matrix to ionize TFV-DP, generation of fragment ions of TFV-DP was investigated as with TFV (Figure 3). At 15% NCE, no fragmentation of TFV-DP (m/z = 446.00281) was observed. Increasing NCE to 20% yielded fragmentation of TFV-DP, producing a fragment ions at m/z = 348.02608 from the parent ion at m/z = 446.00249. NCE at 22% generated the highest abundance of the fragment ion at m/z = 348.02379 from the parent TFV ion at m/z = 445.99881. The m/z = 445.99881 is within the mass tolerance window of 5 ppm. NCE higher than 22% was found to eradicate parent ion TFV-DP signal.

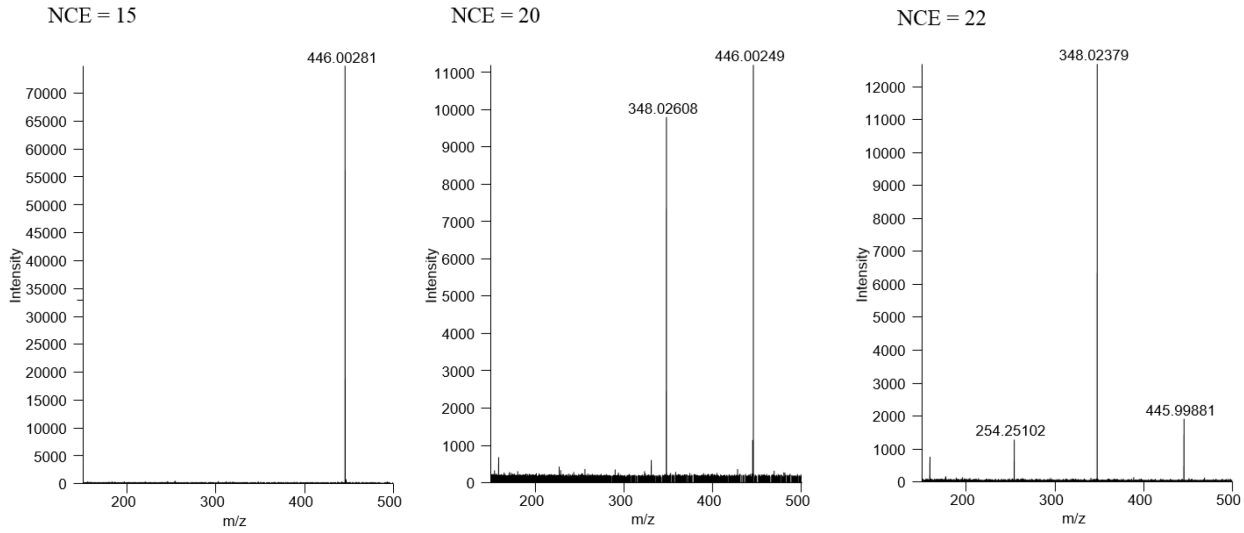

Figure 4. TFV-DP fragmentation at normalized collision energies (NCE) of 15%, 20%, and 22%. No fragmentation of TFV-DP is observed at 15% NCE. The NCE at 22% yielded the greatest intensity for the fragment ion of TFV-DP using the transition at m/z = 445.99881→348.02379.

Spatial distribution of TFV-DP on a steel target plate is presented using MALDI-MS imaging.

Detection methods of TFV-DP, the active anabolite of TFV, were applied toward investigating TFV-DP distribution on a steel target plate. A TFV-DP standard solution at 10 µM final concentration mixed in a 1:1 ratio with 9AA matrix can be visualized when using the extracted fragment ion at m/z = 348 (Figure 3). Heterogeneous TFV-DP distribution was observed in the spotted standard, with TFV-DP unable to be detected uniformly throughout the sample. Water was combined in a 1:1 ratio with 9AA matrix solution as a negative control. No spectra collected in the water sample had a peak at m/z = 348.02379 ± 5 ppm. For TFV-DP detection, the highest spotted concentration was 10 µM and the lowest detectable spotted concentration at 78 nM.

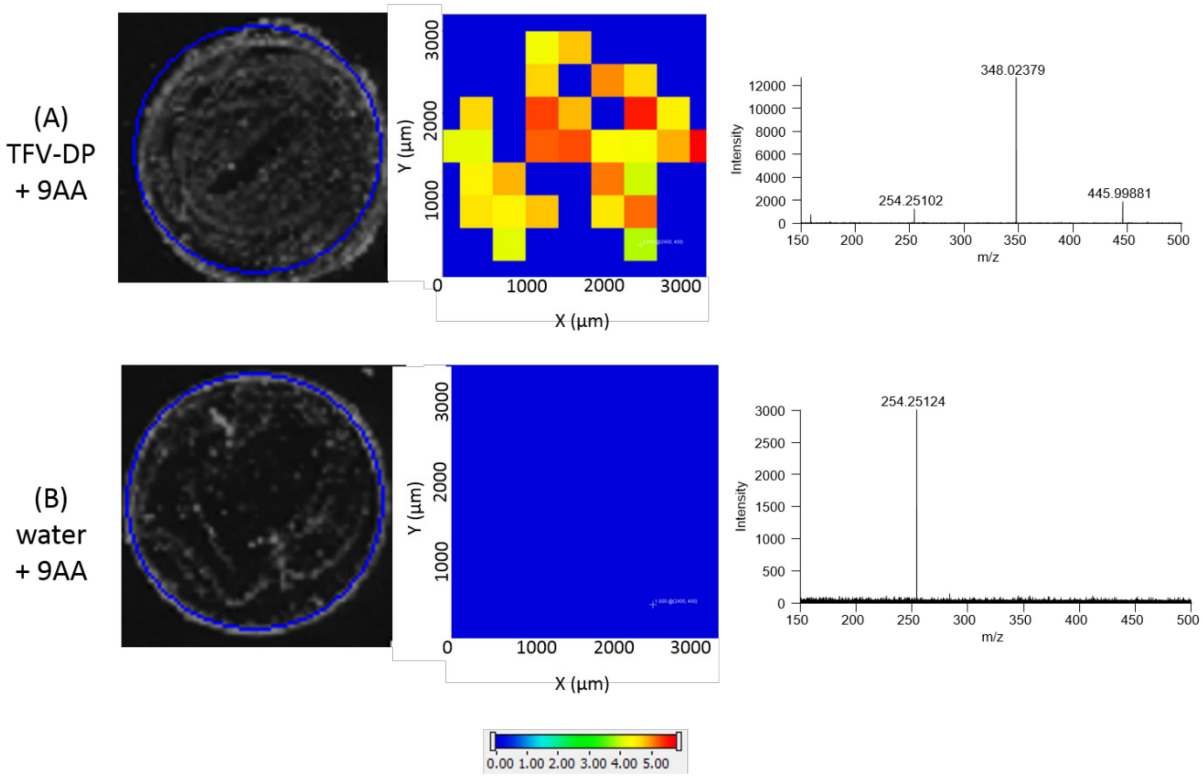

Figure 5. TFV-DP drug standard on a steel target plate visualized using extracted ion MS/MS data. (A) TFV-DP drug standard is combined with 9AA matrix solution in a 1:1 ratio, with 1.5 µL of the resulting solution pipetted onto a steel target plate. Localization of TFV-DP using the fragment ion at m/z = 348.02379 ± 5 ppm is demonstrated in the middle panel. The spectrum displayed is an average spectra across the entire spot scanned by the laser. (B) Water is added in 1:1 ratio to 9AA matrix solution. Extraction of the TFV-DP fragment ion at m/z = 348.02379 ± 5 ppm yields no peaks from the dataset collected for this spot.

Discussion

This method presents a technique to examine TFV and its pharmacologically active anabolite TFV-DP. We are able to establish imaging of TFV using CHCA matrix and detection using a MALDI source coupled to an LTQ-Orbitrap mass spectrometer. Further, TFV-DP has not previously been reported to be detected using MALDI. We employ the matrix 9AA in negative ion mode to demonstrate detection of TFV-DP using MALDI-MS.

In determining which matrices to test in order to best detect TFV, previous studies examining TFV using MALDI as well as the structure of TFV were taken into account. The

matrix CHCA was used to detect TFV in plasma in a previous study [16]. TFV mimics the structure of adenine monophosphate, rendering it a nucleotide monophosphate analog. The matrix 3HPA has been reported to ionize oligonucleotides [17] using time-of-flight mass spectrometry. Ionization of oligonucleotides by 3HPA made it a promising candidate for ionizing TFV, a nucleotide monophosphate analog. Along with CHCA and 3HPA, the matrix DHB was also tested. DHB is a widely used matrix for ionizing small molecules such as peptides and lipids [18, 19], however given the common ionization efficiency of DHB it was still investigated as a matrix for ionizing TFV. Ultimately however, CHCA proved to be the most suitable matrix for detecting TFV using the method we developed.

No prior studies reporting detection of TFV-DP by MALDI have been reported. However, there have been studies reporting nucleotide triphosphate and nucleotide triphosphate analog detection using MALDI coupled to either time-of-flight mass spectrometers or fourier transform ion cyclotron resonance mass spectrometers [20, 21, 22]. One study reported the detection of azetidine-triphosphate, another HIV reverse transcriptase inhibitor using a comatrix of anthranilic acid and nicotinic acid [20] in PBMC. Other studies have reported detection and subsequent imaging of the endogenous nucleotide triphosphate adenosine triphosphate (ATP) in rat brain and in kidney [21, 22]. To detect ATP, the matrix 9AA was used in both studies. The structural similarity between ATP and TFV-DP made the matrix 9AA a good candidate for allowing for TFV-DP detection. This similarity is likely also what made 9AA the most sensitive matrix for TFV-DP detection of the matrices tested.

Along with optimizing the detection of TFV and TFV-DP using an LTQ Orbitrap, imaging these two compounds allows for application of this technique toward investigating the spatial distribution of these drugs in tissue. This work is the first to report molecular imaging of TFV and TFV-DP using MALDI-IMS with an LTQ-Orbitrap XL mass spectrometer. Taken with the drug distribution information that MALDI yields, the high resolution accurate mass

capabilities provided by the Orbitrap aid in providing accurate mass measurements of TFV and TFV-DP.

Future experiments applying this technique can address the compartment specificity of TFV activation in the HIV susceptible PBMC, vaginal, and colorectal tissues. These studies would provide spatial distribution information regarding drug localization and activation. Examining the distribution of TFV would be able to direct PrEP studies regarding the penetration of TFV in tissue. This information would build upon current investigations on the effectiveness of PrEP. In the case of imaging TFV-DP in tissue, this would yield information regarding the activation of TFV. Taken together, the spatial distribution information of both TFV and TFV-DP in tissue that this method could provide would yield a deeper understanding of PrEP efficacy unattainable from measuring drug levels in plasma or homogenates alone.

References

[1] B. S. R. K. C. B. N. F. A. Robbins, "Anti-human immunodeficiency virus activity and cellular metabolism of a potential prodrug of the acyclic nucleoside phosphonate 9-R-(2-phosphonomethoxypropyl)adenine (PMPA).," *Antimicrobial Agents and Chemotherapy,* vol. 42, no. 3, pp. 612-617, 1998.

[2] R. Schooley, "Tenofovir DF in antiretroviral-experienced patients: results from a 48-week, randomized, double-blind study," *AIDS,* vol. 16, no. 9, pp. 1257-1263, 2002.

[3] U. F. a. D. Administration, "FDA Approves First Medication to Reduce HIV Risk.," FDA Consumer Health Information, 2012.

[4] R. Grant, L. KR, A. PL and e. al., "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men.," *N. Engl. J. Med.,* vol. 363, no. 27, pp. 2587-2599, 2010.

[5] J. e. a. Marrazzo, "Pre-exposure prophylaxis for HIV in women: daily oral tenofovir, oral tenofovir/emtricitabine, or vaginal tenofovir gel in the VOICE study (MTN 003).","in *0th Conference on Retroviruses and Opportunistic infections.* , Atlanta, GA., 2013.

[6] C. B. G. V. H. C. J. J. e. a. Hendrix CW, "MTN-001: Randomized Pharmacokinetic Cross-Over Study Comparing Tenofovir Vaginal Gel and Oral Tablets in Vaginal Tissue and Other Compartments.," *PLoS ONE.,* vol. 8, no. 1, 2013.

[7] T. E. H. C. ,. B. N. Lade JM, "Discovery of Genetic Variants of the Kinases That Activate Tenofovir in a Compartment-specific Manner," *EBioMedicine,* vol. 2, no. 9, p. 1145–1152, 2015.

[8] K. P. H. K. E. J. A. e. a. Patterson, "Penetration of Tenofovir and Emtricitabine in Mucosal Tissues: Implications for Prevention of HIV-1 Transmission," *Science Translational Medicine,* vol. 3, no. 112, 2011.

[9] S. S. M. T. G. P. L. B. I. C. A. H. L. G. S. M. F. M. B. L. C. C. F. C. N. M. L. M. P. L. Julien Fonsart, "Single-dose pharmacokinetics and pharmacodynamics of oral tenofovir and emtricitabine in blood, saliva and rectal tissue: a sub-study of the ANRS IPERGAY trial," *J. Antimicrob. Chemother. ,* 2016.

[10] Y.-J. C. P. L. S. R. G. L. S. R. T. S. N. J. R. A. S. E. R. B. E. J. F. a. C. W. H. Nicolette A. Louissaint, "Single Dose Pharmacokinetics of Oral Tenofovir in Plasma, Peripheral Blood Mononuclear Cells, Colonic Tissue, and Vaginal Tissue," *AIDS Research and Human Retroviruses,* vol. 29, no. 11, pp. 1443-1450, 2013.

[11] A. S. M. S. a. B. P. Eric G. Solon, "Autoradiography, MALDI-MS, and SIMS-MS Imaging in Pharmaceutical Discovery and Development," *The AAPS Journal,* vol. 12, no. 1, pp. 11-26, 2012.

[12] A. H. C. McEwen, "Quantitative whole-body autoradiography: past, present and future," *Bioanalysis,* vol. 7, no. 5, pp. 557-568, 2015.

[13] M. C. R. Reyzer, "MALDI-MS-based imaging of small molecules and proteins in tissues.," *Curr Opin Chem Biol. ,* vol. 11, no. 1, pp. 29-35, 2007.

[14] V. K. H. B. R. V. G. S. S. H. Kerstin Strupat, "MALDI Produced Ions Inspected with a Linear Ion Trap-Orbitrap Hybrid Mass Analyzer," *Journal of the American Society for Mass Spectrometry,* vol. 20, no. 8, pp. 1451-1463, 2009.

[15] G. R. T. B. C. T. C. S. A. D. M. K. D. C. M. Jeremy A. Barry, "Mapping Antiretroviral Drugs in Tissue by IR-MALDESI MSI Coupled to the Q Exactive and Comparison with LC-MS/MS SRM Assay," *Journal of The American Society for Mass Spectrometry,* vol. 25, no. 12, pp. 2038-2047, 2014.

[16] v. K. J. S. R. v. d. E. M. G. R. L. T. Meesters RJ, "Determination of the antiretroviral drug tenofovir in plasma from HIV-infected adults by ultrafast isotope dilution MALDI-triple quadrupole tandem mass spectrometry.," *J Mass Spectrom,* vol. 46, no. 3, pp. 282-289, 2011.

[17] S. T. B. C. Wu K.J., " Time-of-flight mass spectrometry of underivatired single-stranded DNA oligomers by matrix-assisted laser desorption," *Anal.Chem.,* vol. 66, p. 1637–1645., 1994.

[18] M. K. F. H. K. Strupat, "2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry," *International Journal of Mass Spectrometry and Ion Processes,* vol. 111, pp. 89-102, 1911.

[19] S. B. Y. H. Sophie Bourciera, "Ionization of 2,5-dihydroxybenzoic acid (DHB) matrix-assisted laser desorption ionization experiments and theoretical study," *International Journal of Mass Spectrometry,* Vols. 210-211, pp. 59-69, 2001.

[20] F. P. H. V. v. R. A. H. N. d. G. R. L. T. van Kampen JJ1, "A new method for analysis of AZT-triphosphate and nucleotide-triphosphates.," *Biochem Biophys Res Commun.,* vol. 315, no. 1, pp. 151-159, 2004.

[21] T. D. B. A. L. O. Benabdellah F, "In situ primary metabolites localization on a rat brain section by chemical mass spectrometry imaging.," *Anal Chem,* vol. 81, no. 13, pp. 5557-5560, 2009.

[22] H. C. H. G. D. M. D.-S. M. D. A. S. L. P. S. S. J. D. P. S. K. Miyamoto S, "Mass Spectrometry Imaging Reveals Elevated Glomerular ATP/AMP in Diabetes/obesity and Identifies Sphingomyelin as a Possible Mediator.," *EBioMedicine.,* vol. 7, pp. 121-134, 2016.

[23] M. Thigpen, K. PM, P. LA, t. T. S. Group and e. al., "Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana.," *N. Engl. J. Med.,* vol. 367, pp. 423-434, 2012.

[24] E. H. S. O. S. K. M. L. A. Sabini, "Structural Basis for Substrate Promiscuity of dCK," *Journal of Molecular Biology,* vol. 378, no. 3, pp. 607-621, 2008.

[25] J. Baeten, "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women.," *N Engl J Med. ,* vol. 367, no. 5, pp. 399-410. , 2012.

[26] J. Baeten, D. Donnell, P. Ndase and t. P. P. S. Team, "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women.," *N. Engl. J. Med,* vol. 367, pp. 399-410, 2012.

Chapter 5: Conclusions

Comprising the only FDA-approved HIV pre-exposure prophylaxis regimen Truvada ®, the nucleoside/nucleotide reverse transcriptase inhibitors tenofovir and emtricitabine require further studies regarding their activation. This is especially necessary since no clinical studies reporting full protection against HIV infection have yet been recorded. Participant adherence to the drug regimen has been shown to play a major role in this gap in protection, however, taking adherence into account has been able to yield 92% protection against HIV infection at most [1]. The studies carried out here begin to address possible sources of variation that could affect PrEP activation.

In looking for genetic variants in the genes that encode the kinases adenylate kinase 2, creatine kinase, muscle, pyruvate kinase, muscle, and pyruvate kinase, liver and red blood cell, we build upon previous studies identifying the phosphorylation scheme of tenofovir [2]. The genetic variants detected in the studies in Chapters 2 and 3, along with their predicted functional impacts, provide a starting point for examining sources of differences in inter-individual tenofovir phosphorylation. Depending on their functional impact, these genetic variants could negatively impact tenofovir phosphorylation, leading to compromised protection from HIV infection.

Along with discovering genetic variants in kinases that play a role in tenofovir phosphorylation, the emtricitabine phosphorylation pathway has also been proposed. In Chapter 3, we put forward a phosphorylation scheme for emtricitabine activation. Genetic variants in thymidine kinase 1, deoxycytidine kinase, cytidine monophosphate kinase 1, and phosphoglycerate kinase 1 have been discovered, but their predicted functional impact must yet be interrogated using *in vitro* studies.

Tantamount to the identification of tenofovir activating kinases is the discovery of their compartment-specific manner of tenofovir phosphorylation. While adenylate kinase 2 was found to phosphorylate tenofovir to tenofovir-monophosphate across peripheral blood mononuclear cells, colorectal tissue, and vaginal tissue, creatine kinase, muscle was demonstrated to phosphorylate tenofovir-monophosphate to tenofovir-diphosphate in colorectal tissue but not in vaginal tissue or peripheral blood mononuclear cells. The kinases pyruvate kinase, muscle and pyruvate kinase, liver and red blood cell were found to generate tenofovir-diphosphate in vaginal tissue and peripheral blood mononuclear cells but not in colorectal tissue. This compartment specificity demonstrates a need for the spatial distribution information of tenofovir and tenofovir activation, addressed by the method developed in Chapter 4. Using matrix-assisted laser desorption/ionization coupled to mass spectrometry, this method is able to preserve distribution information by analyzing tissue slices.

Taken together, these pharmacogenetic and imaging approaches toward understanding tenofovir and emtricitabine activation provide sources of possible differences in HIV PrEP trial outcomes. These data begin to provide a shift in the current understanding of HIV drug metabolism by demonstrating drug activation in colorectal and vaginal tissues, possible sources of HIV transmission. These studies can be used to inform future PrEP clinical trials, working toward full protection from HIV infection.

References

1. R. Grant, L. KR, A. PL and e. al., "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men.," *N. Engl. J. Med.,* vol. 363, no. 27, pp. 2587-2599, 2010.

2. Lade, J.M., To, E.E., Hendrix, C.W., Bumpus, N.N., 2015. Discovery of Genetic Variants of the Kinases That Activate Tenofovir in a Compartment-specific Manner. EBioMedicine. 2 (9), 1145–1152.

Dominique Beltran Figueroa
dfiguer6@jhmi.edu
(317) 332 – 4224
Born July 17th, 1987, Quezon City, Philippines

EDUCATION

2012 - present **Johns Hopkins University School of Medicine,** Baltimore, MD
Ph.D., Pharmacology and Molecular Sciences, December 2016

Thesis Title: *Towards Understanding Tenofovir Disposition: MALDI-MS Imaging and Pharmacogenetics*
Thesis Committee: Dr. James Stivers (chair), Dr. Amanda Brown, Dr. Namandje Bumpus, Dr. Craig Hendrix

2005 – 2009 **Stanford University**, Stanford, CA
Bachelor of Sciences, Biology

2005 - 2003 **The American International School-Salzburg**, Salzburg, Austria

RESEARCH EXPERIENCE

Johns Hopkins University Pharmacology and Molecular Sciences, Bumpus Lab.
(07/2014 – present) Activation and distribution of tenofovir as HIV pre-exposure prophylaxis via MALDI-imaging, LC/MS assay development, protein purification and activity assays, pharmacogenomic screens and clinical trial development.

- Designed and executed novel mass spectrometry-based methodology visualizing anti-HIV drug and drug activation in clinical samples such as vaginal tissue, colorectal tissue, and plasma, contributing to award of $50,000 Hopkins Center for AIDS Research Innovation grant

- Genotyped the largest number of individuals (996) for sources of pharmacokinetic variation in HIV protection efficacy

- Discovered 103 previously unreported genetic variants demonstrated to potentially play a role in clinical anti-HIV drug activation

Publications and Works in Progress

- Figueroa DB, Bumpus NN. "Discovery of genetic variants of the kinases that activate tenofovir among individuals from the United States, Thailand and South Africa." 2016. (Manuscript in preparation).

- Figueroa DB, Bumpus NN. "Investigation of genetic variants discovered in the kinases involved in tenofovir and emtricitabine phosphorylation." 2016. (Manuscript in preparation).

- Figueroa DB, Lade, JM, Bumpus NN. "A new method for imaging tenofovir and tenofovir diphosphate in tissue using matrix assisted laser desorption ionization mass spectrometry." 2016. (Manuscript in preparation).

Stanford University Chemical and Systems Biology, Elias Lab. (01/2010 – 07/2012) Support PI's research, optimizing and maintaining quality control of lab liquid chromatography/mass spectrometry systems.

- Enhance utility of collected proteomic data for collaborators by implementing instrument input improvements
- Designed and established robust, sustainable lab liquid chromatography/mass spectrometry system quality control protocols
- Documented quality control findings in order to ensure excellent instrument condition
- Guided experimental design, sample preparation and analysis for 8 institutional collaborators, resulting in 4 publications
- Inaugurated guidelines for laboratories to streamline and develop institution-wide scientific partnerships

Thrive Research. (03/2009 – 08/2009) Development and efficacy analysis of computer-based health education and improvement programs targeted toward youth and college students, website content management.

Stanford University Medical Center, Obstetrics and Gynecology. (06/2008 - 09/2008) Observe mouse surgeries in preparation for in situ hybridization. Carry out immunocytochemistry protocol in order to localize proteins present in murine placenta.

Stanford University Medical Center Orthopedics, Stanford University Mechanical Engineering, Biomotion Lab. (04/2007 - 09/2007) Process and analyze serum samples using ELISA technique to test varying levels of MMP-9 in osteoarthritic and non-osteoarthritic patients.

Stanford University Medical Center Pathology, Proteomics Integrative Research Facility. (09/2006 - 04/2007) Process and analyze tissue samples for protein identification using gel electrophoresis and MALDI/TOF, mass spectrometry instrumentation.

PROFESSIONAL MEMBERSHIPS

American Society for Pharmacology and Experimental Therapeutics (ASPET), (01/2015 - present)

PRESENTATIONS

Figueroa, DB., Lade, JM., Bumpus, NN. Development of A Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Method for Imaging Tenofovir and Tenofovir-Diphosphate. Oral and poster presentation to be presented at The 21st International Symposium on Microsomes and Drug Oxidations, Davis, CA. October 2016.

MASS SPECTROMETRY AND CLINICAL SAMPLE HANDLING SKILLS

Mass Spectrometry-Based Proteomics: Processing of protein samples for mass spectrometry applications, including in-solution and in-gel digestion, peptide de-salting cleanup, chemical labeling of peptide samples, nanoflow HPLC system troubleshooting, Orbitrap instrument maintenance and quality control, expertise with Thermo instruments and Xcalibur and Proteome Discoverer software.

Mass Spectrometry for Small Molecules: Assay development for small molecules including compound detection optimization in Thermo triple quadrupole mass spectrometer, HPLC optimization for separation of compounds, management, quality control, and maintenance of Thermo triple quadrupole mass spectrometer along with regular cleaning and calibration of instrument.

Mass Spectrometry Imaging: Tissue sample preparation optimization for detection of compounds of interest including harvest, freezing, slicing and matrix selection and application on tissue using HTX TM-sprayer, compound detection optimization in Thermo MALDI-MS system.

Mass Spectrometers Previously Utilized: Thermo LTQ, Thermo LTQ-Orbitrap, Thermo TSQ Vantage, Thermo MALDI-Orbitrap XL, Thermo Q Exactive

Pharmacokinetic and Pharmacodynamic Analysis: Familiarity with fundamental PK-PD parameters, fluency in modeling PK-PD parameters and simulation of clinical studies using WinNonLin.

Next Generation Sequencing: Genomic DNA extraction from dried blood spots, whole blood, and tissue, genomic DNA quality control and quantification using Qubit (fluorimetry) and Take3 (spectrophotometry), custom amplicon design, Illumina TruSeq Custom Amplicon library preparation and quality control, analysis of data using Illumina MiSeq Reporter and Illumina VariantStudio

WORK EXPERIENCE

Life Science Research Assistant, Elias Lab, Chemical and Systems Biology, Stanford University School of Medicine (01/2010 - 08/2012) Maintain instrumentation and collaboration with other labs, both institution-wide and in industrial settings, interpret and analyze results, work with collaborators to execute and optimize mass spectrometry-centered protocols, stay updated with current literature to expand knowledge of new techniques and advanced in related research, carry out general laboratory maintenance and support.

Operations Associate, DocSpot. (08/2009 – 01/2010) Consumer-oriented early stage health-care startup aiming to connect patients with positively reviewed healthcare providers. Determine product development and implementation, oversee company branding, optimize consumer compatibility to product, develop and refine language on website. Authored protocol researching and compiling provider information across 12 metropolitan centers nation-wide. Compiled and organized database of over 10,000 physicians eligible for patient review. Managed company branding, developed logo still in use at Docspot today.

Student Assistant, Falconer Biology Library, Stanford University. (09/2008 - 06/2009) Shelve books, organize library print periodicals, implement electronic periodical database, assist patrons with general library services.

Mentor, High School Summer College Program – Stanford University (06/2006-08/2006) Work alongside 23 undergraduates to plan educational, social, and community service activities for high school students, provide guidance and advice during students' on-campus stay while living in their dormitory setting.

LEADERSHIP AND ADMINISTRATIVE EXPERIENCE

2015 – present **Pharmacology Student Initiative,** Johns Hopkins Pharmacology. Founding member. Facilitate building a stronger network within the Johns Hopkins School of Medicine Pharmacology department on an between both students and faculty by planning and initiating events to foster departmental camaraderie and awareness.

- Instituted department-wide network among faculty, students, and staff, fostering mentoring and scientific collaboration
- Secured inaugural $3,000 budget from department chair toward multiple organization-led event series

- Led and organized annual new student recruiting efforts, resulting in 60% confirmation of desired applicants within 5 days of admission offer

- Improve communication and sustain community engagement among faculty, students, and staff by coordinating bimonthly meetings to structure events

2012 - present **Stanford Outreach Volunteer Alumni Link,** Stanford University Undergraduate Admission. Interview and assess high school applicants for University admission in the Baltimore area.

2010 – 2013 **UNICEF Campus Initiative Alumni Association**, United States Fund for UNICEF. National Council Member. Develop alumni association framework and membership guidelines for a campus-based effort on a national level, coordinate volunteers to continue to engage in raising funds and awareness for a global humanitarian relief organization beyond their time at university.

2010 – 2012 **Laboratory Science for Broadened Scientific Inquiry (LABScI)**, Stanford University, Lucille Packard Children's Hospital. Develop and refine science laboratory curriculum for students in a non-traditional education environment, lead high-school and middle-school science laboratory lessons in hospital classroom.

2007 – 2009 **Stanford University Students for UNICEF**, Stanford University. Financial Officer, Founding Member. Apply for funding for student group focused on raising awareness regarding child poverty, education and health issues worldwide. Coordinate campus-wide events to educate, advocate and fundraise for United States Fund for UNICEF.

2006 – 2009 **Promoting Women's Health and Human Rights**, Stanford University. Board Member/Activism Coordinator. Plan events focused on raising women's issues and liberation awareness. Make efforts toward public and campus outreach regarding issues such as AIDS, FGM, violence against women as well as empowerment and campus identity.

2005 – 2009 **Kayumanggi Filipino Dance Troupe,** Stanford University. Corps member (2004-2005, 2007-2008), Soloist (2005-2006), Co-artistic Director (2008-2009). Choreograph and set traditional Filipino folk dances, organize corps members for monthly and spring showcase performances.

2005 – 2009	**Cardinal Ballet Company**, Stanford University. Soloist (2006-2007), Publicity Officer (2007-2009), Corps de Ballet (2007-2009). Organize advertising for group events, work closely as a team with group members during training, retain group focus while having to add to repertoire upon short notice.
2005 – 2008	**Stanford Student Biodesign**, Stanford University. Networking/Communications Chair Member. Search for and connect biotech companies to both graduate and undergraduate students by organizing career fairs and mentoring programs, compile job offerings from periodicals, publish monthly newsletter.
2005-2007	**Pilipino American Student Union**, Stanford University. General core member; freshman intern. Coordinate campus-wide events raising awareness regarding Filipino culture and issues.

AWARDS

Nominated: Amy J. Blue Award (2012)
The President's Volunteer Service Award: Gold Level (2011-2012)
The President's Volunteer Service Award: Silver Level (2010-2011)
European Council of International Schools Award for International Understanding (2005)